VOL. 2, NO. 3, AUGUST, 1960

human factors

NOV 29

THE JOURNAL OF THE HUMAN FACTORS SOCIETY

Contents

THRESHOLD AND RATE OF DARK ADAPTATION AS FUNCTIONS OF AGE AND TIME 109
Richard G. Domey, Ross A. McFarland and Ernest Chadwick

TARGET SIZE AND VISUAL RECOGNITION 120
William C. Steedman and Charles A. Baker

A LABORATORY EVALUATION OF THE EFFECTS OF ELECTRONIC COUNTERMEASURES ON SYSTEM PERFORMANCE 128
William D. Hitt and Horace W. Ray

A SYSTEM MEASURE OF DRIVER PERFORMANCE 136
David B. Learner

A RE-EVALUATION OF MACHINE-GENERATED ABSTRACTS 141
Albert Resnick and Terry R. Savage

DYNAMIC ANTHROPOMETRY OF WORKING POSITIONS 147
H. T. E. Hertzberg

VOLUMETRIC REQUIREMENTS FOR HAND TOOL USAGE 156
Paul T. Baker, James M. McKendry and George Grant

ACOUSTIC NOISE, AN INFORMATIONAL DEFINITION 163
Alan A. Burrows

HUMAN ENGINEERING—1911 STYLE 169
Sherman Ross and Andrew A. Aines

THE AUTHORS 171

OHIO STATE UNIVERSITY
EDUCATION LIBRARY

PERGAMON PRESS NEW YORK • LONDON • PARIS • LOS ANGELES

OFFICERS AND EXECUTIVE COUNCIL

The Human Factors Society is a non-profit California Corporation. The officers and members of the Executive Council are:

PRESIDENT: Arnold M. Small, 1960
PRESIDENT-ELECT: Stanley N. Roscoe, 1960
VICE-PRESIDENT: Smith W. Ames, 1960
PAST-PRESIDENT: Renato Contini, 1960
SECRETARY: Paul G. Cheatham, 1960
TREASURER: Donald W. Conover, 1960
EDITOR-IN-CHIEF: Stanley Lippert, 1959–1961

EXECUTIVE COUNCIL: Charles I. Barron, 1960–1962; Paul M. Fitts, 1959–1960; Jack A. Kraft, 1959–1960; Max W. Lund, 1959–1961; John Lyman, 1960–1962; Ross A. McFarland, 1959–1961

EDITORIAL BOARD: Walter P. Grether; Blaine H. Levedahl; J. C. R. Licklider; Stanley Lippert; John Lyman; John P. Meehan; Allison K. Simons; Ezra V. Saul; Neil D. Warren; Joseph W. Wulfeck

HISTORIAN: Donald T. Hanifan

Address general correspondence to Human Factors Society, P.O. Box 263, Arlington 10, Virginia, Attention: **Paul G. Cheatham,** Secretary, for Society business including payment of dues.

Address matters pertaining to the Journal or Bulletin to **Human Factors,** Biotechnology Laboratory, Department of Engineering, University of California, Los Angeles 24, California, Attention:

John Lyman, Managing Editor, for Journal or Bulletin Subscriptions or other publication business.

Stanley Lippert, Editor-in-Chief, for submission of Journal articles and letters to the editor.

Hilde Groth, Bulletin Editor, for Bulletin news.

COPYRIGHT © 1960
by the HUMAN FACTORS SOCIETY

PUBLISHED BY
PERGAMON PRESS, 122 E. 55th St., New York 22, N.Y.
Headington Hill Hall, Oxford, England.

Printed in Great Britain.

SUBSCRIPTION PRICE: MEMBERSHIP in Society includes subscription to both publications.

Journal: NON-MEMBERS SUBSCRIPTIONS TO BE SENT DIRECT TO THE PUBLISHER, PERGAMON PRESS. Commencing with Volume 2 the subscription price will be:
$10.00 (£3 10s.) per annum, for individuals who certify that the journal is for their personal use. $20.00 (£7) per annum, for libraries, government establishments and research institutions.

Bulletin: $5.00 per year. Orders to be sent to John Lyman.

Threshold and Rate of Dark Adaptation as Functions of Age and Time[1]

RICHARD G. DOMEY, ROSS A. McFARLAND and ERNEST CHADWICK, Harvard School of Public Health, Boston, Massachusetts.

In order to describe one family of dark adaptation curves obtained from an age sample ranging from 16 through 89 years, a mathematical model of the form

$$\log y = G^{(Ct)^{f_1(A)}}$$

was derived.

It was found that the model could be generalized. In addition it was concluded that threshold of dark adaptation as a function of time was lawfully related to chronological age.

Rate of adaptation was determined by differentiating the equation at time 30 sec and time 6 min as follows:

$$R = [G^{(Ct)^{f_1(A)}}][(Ct)^{f_1(A)}/t]f_1(A)\ln G$$

The results showed that rate of dark adaptation was a curvilinear function of age.

These findings are consistent with the hypotheses that dark adaptation threshold and rate of dark adaptation depend upon basic underlying physiological processes that change with age.

DARK ADAPTATION AS A FUNCTION OF AGE AND TIME

The purpose of this paper is to present a mathematical derivation[2] of a model for representing dark adaptation as a function of age and time. In an alternative statistical study of the relationship of dark adaptation and age, McFarland and Fisher[2] stated that the dark adaptation curve itself could be described as inversely logarithmic, and therefore represented by the general equation:

$$y = 10^{a+bx}+C \qquad (1)$$

where y = $\mu\mu$l luminance

a = initial level of dark adaptation

b = drop time[3]

x = time

C = the asymptote of the curve.

However, McFarland[2] shows that it is also possible to compute C by the formula:

$$C = \frac{P_1 \cdot P_2 - P_3^2}{P_1 + P_2 - 2(P_3)} \qquad (2)$$

where P_1 = a given value of the curve early in time, say, the 6th min,

P_2 = a value on the curve late in time, say, the 30th min,

P_3 = a value on the curve midway between P_1 and P_2,

C = the asymptote of the curve.

This method of calculating C, the asymptote of the rod curve, is generally applicable to calculating the asymptote of the cone curve as well.

Hammond and Lee[1] state that the dark adaptation curve can be represented by:

$$\log I = a + b/t^2 \qquad (3)$$

where a = a constant, the general level of the curve at the asymptote,

b = a constant, the "rate" of adaptation,

t = the time in minutes from the cessation of the pre-adaptation light stimulus.

In previous investigations the general procedure was to fit the data by using various formulae, and then to identify such parameters as the curve intercept, cone–rod curve intersection, drop rate, asymptote of the cone curve and the asymptote of the rod curve. But dark adaptation is a phenomenon that is known to

1 Sponsored by the Commission on Accidental Trauma of the Armed Forces Epidemiological Board and supported in part by the Office of the Surgeon General Department of the Army.

2 Since the model was derived from experimental data it is to be interpreted as an empirical one.

3 Drop time is generally taken to mean the amount of time required to reach some predetermined level of dark adaptation.

vary relative to a number of conditions of which age is one. It is now certain that as age increases the curve is displaced upward on the y axis. At the same time the cone and rod segments of the curve appear to pivot around their individual focal points near the intercept. The pattern of displacement is extremely orderly.

It it quite evident that if the age variable is related to the dark adaptation curve in some lawful manner then this function could be described mathematically. If this is so then it should be possible to integrate the age curve family. However, this has not been done.

Instead, at this juncture in the history of investigating dark adaptation as a function of age, conventional statistical methods in the study of the age factor were substituted for a more mathematical resolution of the dark adaptation time and age interaction. This approach established beyond a reasonable doubt that dark adaptation becomes a function of age. For instance, in 1955 McFarland and Fisher [2] used formula (2) for obtaining C and found a correlation of 0.895 between age and C. Since the sample of Ss used was composed largely of aircraft pilots who were, by the nature of their occupation, highly selected with respect to visual efficiency, a correlation of this magnitude would not ordinarily have been expected. This is because restricted range of samples tends to lower the magnitude of the conventional product moment r. The results, $r = 0.895$, however, failed to support this inference because this correlation was one of the highest ever discovered among physiological and psychological relationships.

However, the statistical approach did not fulfill the need for a mathematical model of the dark adaptation, time and age relationships. Consequently, new data obtained by McFarland, Domey and Warren were examined, and a search for a general integrative equation was initiated. A model was constructed which made possible the accurate prediction of the mean level of adaptation for any point on the time continuum as a function of age.

METHOD

Subjects

There were 240 male subjects since 30 subjects were drawn from each of 8 decades ranging from the teenage level through the age of 89. Previous studies have shown that there is no significant sex difference in the dark adaptation phenomenon. The total sample of 240 subjects was composed of persons taken from YMCA youth groups, college age students, university faculty, taxi drivers, unemployed persons obtained from the USES Agency and retired men living at home or in private institutions for the aged. All subjects were paid for their services, and subsequent to obtaining the data, over half the subjects in each decade were then offered and given a complete eye examination free of charge.

Apparatus

The instrument used throughout this study was the Hecht–Schlaer Adaptometer which had been rebuilt and recalibrated by the manufacturer, O. C. Rudolph and Sons, Caldwell, New Jersey, Model No. 1, Serial No. 14. The apparatus was housed in one of three dark rooms in each of three different cities, in order to accommodate aged persons for whom traveling was difficult.

Procedure

After each subject was seated in the experimental room his left eye was covered by a patch, and his head was held steady in a head-chin rest. Vision was uncorrected. The lights in the dark room were turned off, and after a lapse of approximately one minute, the retina of the right eye was bleached by exposure to a standard 1600 ml incandescent light source for 3 min. At the end of the pre-test period, the subject was exposed to a red fixation point 7° right of center. The 1° violet test stimulus (53.0 per cent transmission at 405 mμ) was then presented. The duration of each test flash of light was $\frac{1}{5}$ of a second.

The first observation was made within the first 59 seconds after the termination of the

pre-exposure light. Then, beginning with the second observation, one reading was taken every minute for the next 9 min, every 2 min for the next 6 min, and every 3 min for the following 24 min. Therefore, 21 observations were made in 40 min of dark adaptation.

RESULTS

The numerical results are presented in Table I where $\log_{10}$ $\mu\mu$l luminance in the body of the table is given as a function of age and time. Fig. 1 constructed from the data found in Table I shows that the family of dark adaptation curves obtained as a function of age rises in an orderly manner, suggesting that the differences between their terminal points increases geometrically from one age level to the next. Since the curves for the 16–19-year-old group and the 20–29-year-old group appear to be nearly alike they have been combined for the purpose of simplifying the mathematical process which follows, although they are plotted independently in Fig. 1.

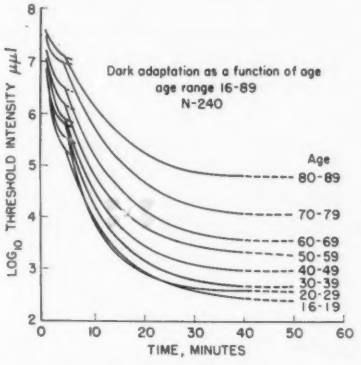

Fig. 1. Conventional dark adaptation curves which show the pronounced influence of age upon threshold of dark adaptation.

AN ANALYTIC EXPRESSION FOR DARK ADAPTATION

For the sake of convenience in examining the level of dark adaptation as a function of age as well as time, the mean data have been replotted as *time* curves as shown in Fig. 2. Here age has been entered on the x axis, and $\log_{10}$ $\mu\mu$l luminance entered on the y axis. Each individual curve in Fig. 2 represents the level of dark adaptation as a function of age at a given time in log units. For instance, the top curve in Fig. 2 was drawn from the plot of the data in the first row of Table I, the next curve from the next row and so on. It can be seen that the slope of the curves becomes greater as a function of age and time.

When plotted in this way the data appear as a family of positive exponential curves. It was

TABLE I

Mean Dark Adaptation as a Function of Age and Time ($\log_{10}$ $\mu\mu$l Luminance)
N = 240 (30 in each age group)

Age/Years:	16–19		20–29		30–39		40–49		50–59		60–69		70–79		80–89	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
0–59 sec	6.93	0.30	6.78	0.41	6.95	0.32	7.03	0.26	7.06	0.26	7.21	0.24	7.49	0.58	7.67	0.13
min																
2	6.42	0.32	6.16	0.46	6.37	0.34	6.62	0.27	6.78	0.27	6.94	0.25	7.16	0.57	7.43	0.14
3	5.93	0.39	5.65	0.48	6.05	0.35	6.29	0.32	6.49	0.31	6.73	0.31	7.06	0.17	7.25	0.15
4	5.66	0.39	5.43	0.51	5.85	0.40	5.98	0.35	6.25	0.40	6.59	0.34	6.95	0.22	7.12	0.16
5	5.58	0.40	5.26	0.60	5.67	0.47	5.86	0.37	6.17	0.42	6.50	0.43	6.94	0.23	7.08	0.17
6	5.36	0.41	5.05	0.67	5.45	0.48	5.61	0.33	5.94	0.46	6.27	0.38	6.70	0.29	6.90	0.21
7	5.08	0.45	4.83	0.71	5.17	0.46	5.36	0.34	5.63	0.44	6.02	0.38	6.48	0.29	6.71	0.26
8	4.82	0.46	4.61	0.66	4.86	0.47	5.10	0.38	5.41	0.45	5.76	0.44	6.25	0.32	6.52	0.29
9	4.55	0.52	4.38	0.66	4.84	0.38	4.80	0.38	5.14	0.50	5.51	0.51	5.98	0.43	6.33	0.43
10	4.17	0.48	3.95	0.63	4.17	0.36	4.51	0.39	4.77	0.60	5.24	0.65	5.71	0.44	6.10	0.45
12	3.80	0.40	3.50	0.46	3.87	0.38	4.22	0.34	4.48	0.64	4.99	0.61	5.48	0.46	5.91	0.50
14	3.48	0.43	3.28	0.47	3.62	0.37	3.96	0.35	4.25	0.61	4.78	0.61	5.26	0.45	5.74	0.55
16	3.24	0.41	3.14	0.47	3.44	0.37	3.76	0.31	4.02	0.57	4.50	0.59	5.00	0.48	5.55	0.59
19	3.02	0.35	2.98	0.45	3.23	0.37	3.54	0.31	3.85	0.57	4.27	0.59	4.74	0.48	5.54	0.59
22	2.86	0.35	2.85	0.45	3.04	0.29	3.34	0.29	3.70	0.58	4.05	0.55	4.54	0.47	5.16	0.62
25	2.74	0.32	2.78	0.45	2.93	0.30	3.22	0.29	3.55	0.53	3.88	0.53	4.36	0.46	5.01	0.66
28	2.64	0.30	2.72	0.46	2.87	0.27	3.13	0.30	3.43	0.51	3.74	0.52	4.22	0.46	4.90	0.68
31	2.52	0.30	2.65	0.47	2.81	0.28	3.07	0.31	3.36	0.48	3.67	0.50	4.14	0.47	4.83	0.71
34	2.45	0.23	2.64	0.47	2.77	0.26	3.02	0.29	3.32	0.47	3.63	0.49	4.11	0.47	4.81	0.72
37	2.43	0.17	2.60	0.46	2.76	0.25	3.02	0.29	3.32	0.47	3.62	0.48	4.11	0.48	4.81	0.73
40	2.43	0.17	2.60	0.44	2.76	0.26	3.02	0.29	3.32	0.47	3.62	0.48	4.11	0.48	4.81	0.72

EDUCATION LIBRARY

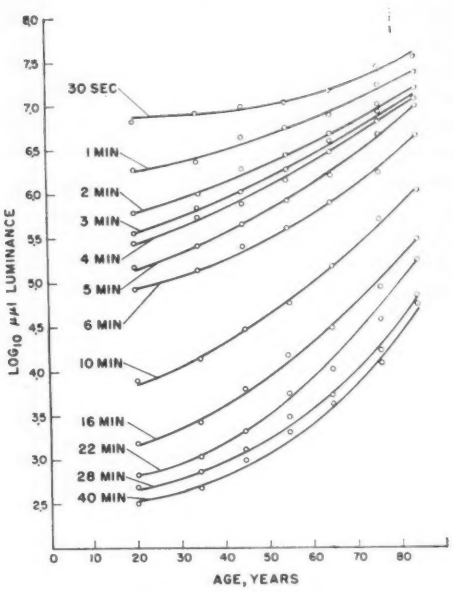

Fig. 2. Dark adaptation curves as a function of time plotted against age.

found that the integration of the age term into the mathematical model of dark adaptation as a function of age and time could be represented as:

$$\log y = G^{(Cf)^{f_1(A)}} \qquad \text{(Model A)}$$

The definition of the parameters will appear in the derivation. It will be seen that the form of the model for the cone family was identical with the form of the model for the rod family.

An examination of the graphic data in Fig. 2 suggests that it would be possible to fit each one of the curves independently. Since the hypothesis was that the family of curves appears to be an orderly function of age, constants common to all the curves were sought through several mathematical techniques. Such constants, if they could be isolated, would contain the key to the integration of the several curves that would permit writing a mathematical expression of the decline of one important human function.

It was eventually discovered that a log log transformation of the original $\log_{10} \mu\mu$l luminance curves shown in Fig. 2 resulted in reducing the distributions from curvilinear form to linear form. The straight line plots of the transformed data are shown in Fig. 3. The curves in Fig. 3 appear as reversed curves because the log log transformation values are negative. Hence the plots here reflect the *sign* of the numbers, not a change in function. Then, by the convenient method of least squares the slope, m, and the constant, k, for each of the seven rod curves were computed. The explanation will be in terms of rod curves. It is understood that the cone curves were treated in exactly the same manner. Now the method for treating all the time–age curves became self-evident. Thus, a plot of the derived slopes, $m_1, m_2 \ldots m_7 \ldots$ was made, as well as a plot of the derived constants, $k_1, k_2 \ldots k_7$. The resulting curves of both the slopes, ms, and the constants, ks, for both cones and rods appear in Figs. 4, 5, 6 and 7. It was found that the first five but not the first six ms and ks could be fitted by the method described.

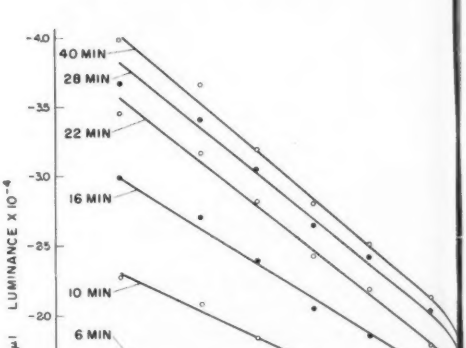

Fig. 3. Triple log transformation of dark adaptation as a function of time plotted against age. The result is a family of straight lines.

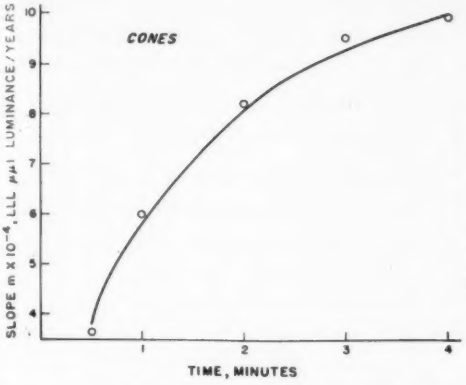

Fig. 4. The values of the first five slopes (cones) shown in Fig. 3 plotted against time.

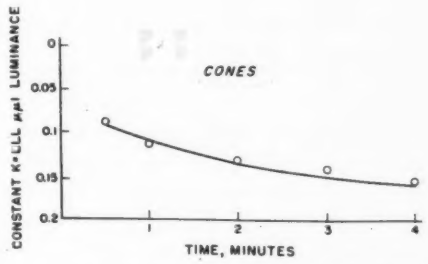

Fig. 5. The values of the constants of the first five slopes (cones) plotted against time.

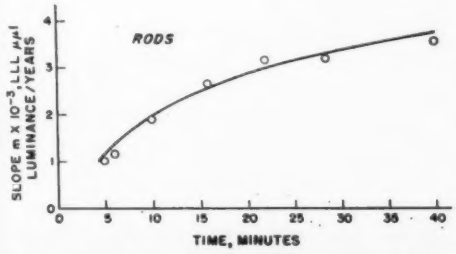

Fig. 6. The values of the last seven slopes (rods) plotted against time.

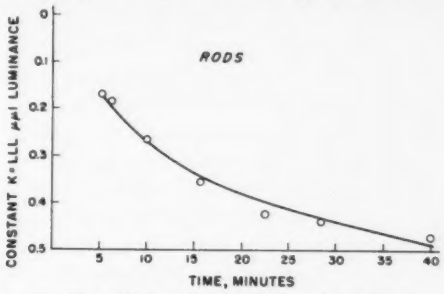

Fig. 7. The values of the constants of the last seven slopes (rods) plotted against time.

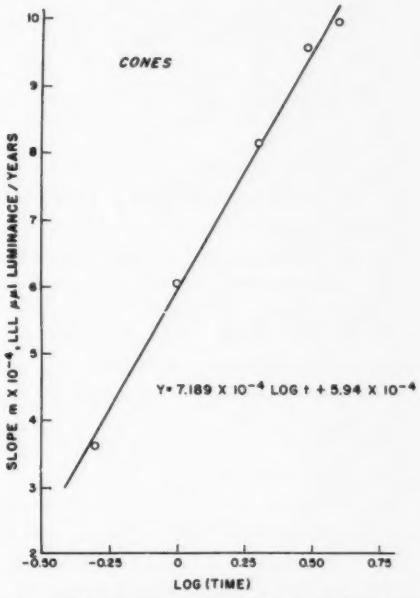

Fig. 8. Triple log transformation of the first five slopes (cones) plotted against log of time.

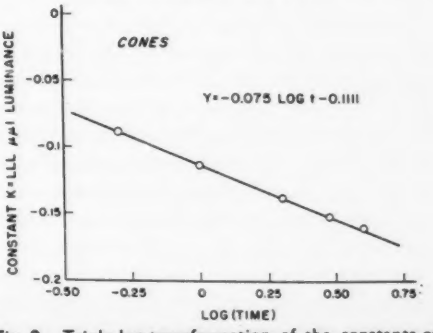

Fig. 9. Triple log transformation of the constants of the first five slopes plotted against log of time.

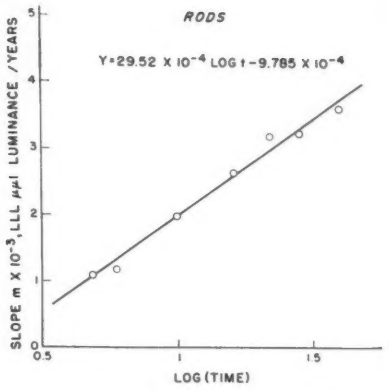

Fig. 10. Triple log transformation of the last seven slopes (rods) plotted against log of time.

It was also found that the last seven but not the last eight ms and ks could be fitted by the method described. Thus, the interval between the slopes and constants at the 4th–5th min defined any cone–rod transition point. It can readily be seen that deriving the formula for these second order curves would provide the constants necessary to the integration of the entire family of time–age curves, and would therefore furnish the means for constructing a mathematical model of dark adaptation as a function of the covariation of time and age.

Since these second order curves were not linear, an appropriate transformation was also sought in order to simplify the mathematical treatment. It was discovered that the m_1, m_2 . . . m_7, and k_1, k_2 . . . k_7 plots could be reduced to linear form by a *log–time* transformation of the illumination parameter in these distributions. The success of this procedure can be clearly seen in Figs. 8, 9, 10 and 11. It followed once again that the application of the method of least squares to these transformed data would produce a slope-of-the-slopes, M_m, along with its own constant K_k. Since the data for cones and rods were plotted independently there were duplicate expressions for both curves. Thus, M_m, M_k, K_m and K_k represent the change in dark adaptation as a function of age and time since they perform the function of constants by tying all the curves together simultaneously. The coefficients for the cone equation are:

M_{mc}, slope of the slopes $= 7.189 \times 10^{-4}$

K_{mc}, constant of the slopes $= 5.94 \times 10^{-4}$

M_{kc}, slope of the constants $= -0.0750$

K_{kc}, constant of the constants $= -0.1111$

$\log t$, the time variable from zero time to infinity

A, the age variable from $x_1 \ldots x_\infty$

The complete explicit expression for the cones is

$$\begin{aligned} &\mathrm{LLL}_{10}\mu\mu\mathrm{l}\,I \\ &= (7.189 \times 10^{-4} \times \log t \\ &\quad + 5.94 \times 10^{-4})\,A - 0.0750 \times \log t - 0.1111 \end{aligned}$$

The rod curves can be represented by

$$\begin{aligned} &\mathrm{LLL}_{10}\mu\mu\mathrm{l}\,I \\ &= (29.52 \times 10 \times \log t \\ &\quad - 0.785 \times 10)\,A - 0.364 \times \log t + 0.08814 \end{aligned}$$

where A represents age, and t represents time.

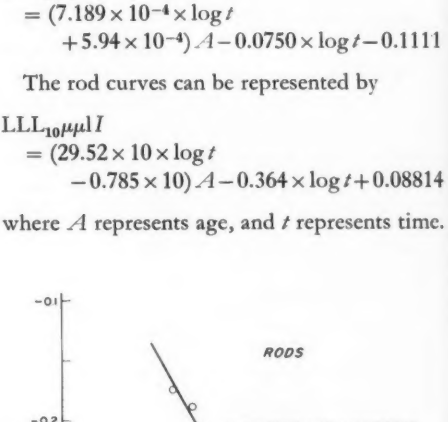

Fig. 11. Triple log transformation of the constants of the last seven slopes (rods) plotted against log of time.

In explicit form the model for the cone curve appears in $\log_{10}$ form as follows:

$$\log_{10}\mu\mu l I = 10^{10^{(5.94\times10^{-4}A-0.1111)}f(7.189\times10^{-4}A-0.0750)}$$

The explicit form of the model for the rod curves is parallel, and is expressed as:

$$\log\mu\mu l I = 10^{10^{(9.785\times10^{-4}A+0.088814)}f(29.52\times10^{-4}A-0.3640)}$$

The explicit form of the model may be reduced to the general form in this way,

$$\log I = G^{(Cf)^{h(A)}} \qquad \text{(Model A)}$$

It can be seen that the determination of the boundaries of the derivation is dependent upon several parameters, (a) level of luminance which cannot be zero, but must reach some least threshold value, (b) time in the dark which is not infinite except theoretically, since the (c) age range within which the dark adaptation phenomenon must exist must be more than zero and less than infinity. For practical and theoretical purposes it may be postulated that the range of luminance within which the model is estimated to be valid would be between approximately 10 and 2 $\log_{10}$ luminance $\mu\mu l$ within an age range between approximately 1 year to 100 years, and a time in the dark period of approximately 60 min. Theoretically the dark adaptation curve is asymptotic. In reality any given dark adaptation curve could not exceed the life span of a subject. And in practical terms more than 98 per cent of the dark adaptation curve is obtained within the first 60 min. Thus, luminance, age and time values could never be zero or negative.

In order to evaluate the derivation two procedures were devised.

Test Procedure 1

In Procedure 1 mathematical Model A was used to reconstruct theoretical dark adaptation curves. Then the original data were compared with the theoretical constructs. Both sets of curves are exhibited in Fig. 12. While the dark adaptation level for the 80–89-year-old group was slightly overestimated, probably because of the concentration of several cataract defects found in persons who were at the far end of the age range, no serious challenge to the efficiency of the derivations was present. The correspondence of the predicted mean curves with the obtained mean curves is unusually close, so close, that one could be substituted for the other

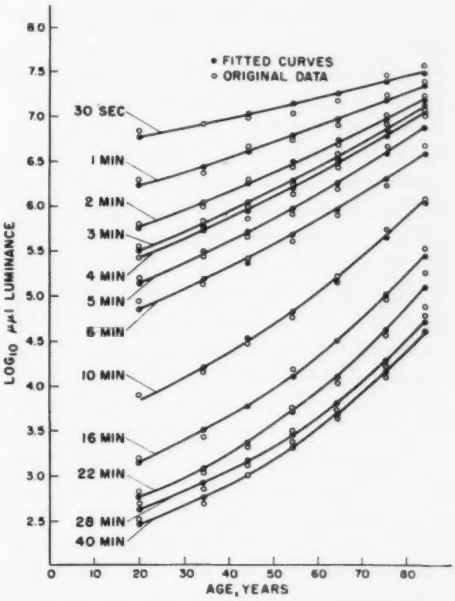

Fig. 12. Comparison of the original data as shown in Fig. 2 with the reconstructed data obtained from Model A. Model A is shown to fit the original data very closely.

without doing violence to the interaction of the age–time co-variables in the dark adaptation phenomenon. For instance, test re-test dark adaptation data usually vary by a displacement of half a log value. Model A fits the present data with far greater exactness.

Since the theoretical curves so closely resemble the original distributions the model was considered valid for the obtained data. What remained was the need to test whether the model could be generalized, that is, extended to

alternative *independent* samples. Test procedure 2 was designed to assist in evaluating this question.

Test Procedure 2

A more rigid test of the generality of the equations would depend upon demonstrating that the model derived from one sample predicted the performance of other independent samples. The present data were treated in the following way to enable the execution of such a test.

Obviously the original data could have been divided into two samples, say, by some unbiased sampling method such as the conventional split-half procedure. Then the equations representing dark adaptation as a function of age and time could have been derived using half the data for all the groups. Finally, these equations could be used to predict those sample data that did not enter into the derivation of the equations. The closeness of fit of the actual curves obtained from the independent sample with the curves predicted by the equations obtained from the data of the criterion group would indicate the degree of coincidence of the mathematical model with the functions it was supposed to represent. Nonetheless, this was not the procedure employed in Test 2. And the reason it was not employed was because a second, more rigorous, method was devised.

The second test procedure was divided into two parts, (1) one part for cone data, and (2) one part for rod data. The reason for treating cones and rods independently was to demonstrate that the stability of the model was not dependent upon the method of sampling.

Test Procedure 2a for Cone Data

The principle of testing the mathematical model by deriving the data from one group and then predicting the performance of a second independent group was modified in the following way. It will be recalled that there were eight age groups. They were assigned to two general classes, a criterion category and a test category. See Table II.

Then using only the data contained within the criterion group, a second model, Model B for cones, was derived independently from Model A. Therefore, there remained three independent test groups the data from which did not enter into the construction of Model B, and thus could in no way have influenced the new Model.

TABLE II

Age Groups for Model B

Criterion category (cones) decades	*Test category (cones) decades* (1)	(2)	(3)
16–29*			
	30–39		
40–49			
		50–59	
60–69			
			70–79
80–89			

* The 16–19, 20–29 age groups were combined, hence the enlarged range, 16–29.

Model B (cones) based upon only those scores inherent in the criterion groups was then compared with the original model, Model A, which was dependent upon all the sampled scores. A comparison of the several coefficients and exponents appears in the explicit formulae below.

Model A (Cones). Original Model Inclusive of All the Data

$$\mathrm{LLL}_{10} I\mu\mu\mathrm{l} = (7.189\times10^{-4}\log t - 5.94\times10^{-4})A - 0.0750\log t - 0.1111$$

Model B (Cones). Second Derivation of the Model Inclusive of Only Those Data Obtained from the Criterion Group

$$\mathrm{LLLL}_{10}\, I\ \mu\mu\mathrm{l} = (7.282\times10^{-4}\log t - 5.73\times10^{-4})A - 0.778\log t - 0.1098$$

Clearly the model for cones remained exactly the same with practically no changes occurring in the coefficients and exponents. In other words Model B can be substituted for Model A.

Test Procedure 2b for Rod Data

In the procedure applied to rod data, the age groups were assigned to two categories, a criterion category and a test category, in the following way as shown in Table III.

TABLE III

Age Groups for Model C

Criterion category decades		*Test category decades*	
	(1)	(2)	(3)
. .	16–29* .	. .	.
30–39 .	. .	. .	.
40–49 .	. .	. .	.
50–59 .	. .	. .	.
60–69 .	. .	. .	.
. .	. .	70–79 .	.
. .	. .	. .	80–89

* The 16–29, 20–29 age groups were combined, hence the enlarged range, 16–29.

Then using only the data contained within the criterion groups, that is, the middle 50 per cent of the age range data, Model C for rods was reconstructed. There remained three independent test groups, all at the extremes of the age range distribution. The age category 16–29 contained 25 per cent of the scores, and each of the age range categories, 70–79, and 80–89 contained 12.50 per cent of the scores respectively. Thus, half the data were used for deriving Model C for rods, and half for test scores.

The comparison of the components of Model A with Model C for rods is as follows:

Model A (Rods). Original Model Inclusive of all the Data

$$\mathrm{LLL}_{10} I\mu\mu\mathrm{l} = (29.52\times 10^{-4}\log t - 9.785 \times 10^{-4})A - 0.3640\log t - 0.8814$$

Model C (Rods). Second Derivation of the Model Inclusive of Only Those Data Obtained from the Criterion Group

$$\mathrm{LLL}_{10} I\mu\mu\mathrm{l} = (28.930\times 10^{-4}\log t - 10.518 \times 10^{-4})A - 0.3665\log t - 0.9487$$

Once again there was no modification of the form of the model, and practically no change in the coefficients and exponents larger than would be expected within rounding error.

It follows that Models B and C are virtually identical to Model A not only in form but in coefficients and exponents as well. Since Model A fits the data with great fidelity so will Models B and C. Hence, Models B and C will predict the scores of their representative test groups which are for the cone model the mean scores of age groups 30–39, 50–59, and 70–79, and which are for the rod model the means of the age groups 16–29, 70–79, and 80–89. Therefore, it has been established that the mathematical

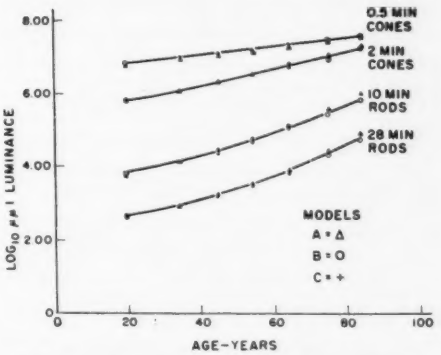

Fig. 13. Comparisons among Models A, B, and C.

derivation of a model of dark adaptation as a function of age and time represents (a) the original data, and (b) predicts the performance of other independent groups of both the same or different age levels. Therefore the model can be generalized.

In order to demonstrate the similarity among Models A, B, C, the curves for all ages at time intervals 30 sec, 2, 10 and 28 min have been calculated using Models A, B, and C independently. Reproducing the entire range of curves for each model would be redundant since any possible differences found among the models would remain constant, or approximately so, and thus would be as easily revealed at one time interval as another. The results appear in Fig. 13.

Since there is now assurance that Model A can be generalized, an analysis of the relationship of rate of dark adaptation in terms of the model cannot be trivial. In order to simplify the mathematical technique in the following analysis, Model A has been used to derive rate of dark adaptation that would appear at the mean ages of 20, 30, 40, 50, 60, 70, 80, 90 and 100 years. The original data through the use of the model have simply been extrapolated to ages 10, 90 and 100 years. Since all the mean ages exist within the boundaries of the model no mathematical rule has been broken when these are substituted for the actual mean ages which are 19.7, 34.2, 44.3, 54.2, 64.2, 75.5, and 83.9.

The equation of the cone curve was differentiated in order to obtain an equation for rate of dark adaptation. Thus,

$$R = [G^{(Ct)^{f_1(A)}}][(Ct)^{f_1(A)}/t]f_1(A)\ln G$$

The constants for the original equation, Model A, were applied to the cone curves and evaluated as follows:

$$Z = \frac{K_{mc}}{M_{mc}} = \frac{5.940\times10^{-4}}{7.189\times10^{-4}} = 0.8263$$

$$B = K_{kc} = \frac{K_{mc}M_{kc}}{M_{mc}} = -0.0481$$

$$D = 10^{B} = 10^{-0.0491} = 0.89309$$

$$C = 10^{Z} = 10^{-0.8263} = 6.7035$$

$$G = 10^{D} = 10^{-0.89309} = 7.8178$$

In an identical manner the constants for the original model, Model A, were applied to the rod curves and evaluated as follows:

$$Z = \frac{K_{mr}}{M_{mr}} = \frac{-9.785\times10^{-4}}{29.52\times10^{-4}} = -0.3315$$

$$B = K_{kr} - \frac{K_{mr}M_{kr}}{M_{mc}} = -0.15$$

$$D = 10^{B} = 10^{-0.15} = 0.7080$$

$$C = 10^{Z} = 10^{-0.3315} = 0.4661$$

$$G = 10^{D} = 10^{+0.7080} = 5.021$$

The data for rate of dark adaptation at the 30th min and the 6th min for mean ages 10, 20 . . . 100 years will be found in Table IV, and in Figs. 14 and 15. It can be seen that when the dark adaptation rate data were plotted as a function of age the resulting distributions were curvilinear. It was discovered that rate appeared to be part of a cosine function when plotted over the independent variable, that is, age. Thus, the age variable was divided by 10 and transformed into angular degrees in order to keep the data within the correct angular range,

TABLE IV

Rate of Dark Adaptation

Mean age	Log_{10} $\mu\mu$l *Luminance/min*	
	Time: 30 sec	Time: 6 min
10	−2.13237	−0.1241
20	−1.95500	−0.1208
30	−1.76747	−0.1166
40	−1.56913	−0.1114
50	−1.35980	−0.1051
60	−1.13860	−0.0973
70	−0.90500	−0.0881
80	−0.65833	−0.0769
90	−0.39790	−0.0636
100	−0.12346	−0.0480

0°–90°, when the data were evaluated. For cone data it was found that an additive constant, 17°, was necessary in the transformation equation. The functions are shown as follows: For the data at the 30th sec the form is:

$$\phi_1 = \frac{(\text{Age})}{10} + 17°,$$

where the constant, 17°, in ϕ_1 was determined empirically and is valid only for the experimental conditions described in this paper. This constant will have to be re-established when the conditions vary.

$$\text{Rate} = a\cos\phi_1 + b$$

where a = a constant, slope

b = a constant, the intercept

For the rate of adaptation of rods as a function of age at 6 min the transformation is:

$$\phi_2 = \frac{\text{Age}}{10}$$

$$\text{Rate} = a\cos\phi_2 + b$$

The curvilinear form of the data and the linear transformation data are shown in Figs. 14 and 15. The success of the transformation is evident. All that remained was the derivation of the equations which would express rate of dark adaptation as a function of age for cone and rods. It was equally clear that the equations for the two linear transformed slopes were the expressions showing the relationship between rate of dark adaptation and age. The slopes were derived by the method of averages in the following way.

Each slope was divided into two segments and the values substituted into the equation $R = ab\cos\phi$. Since, for each curve there were two segments with 5 values in each segment, the result was a total of 10 equations from which, by the addition of five different ones at a time, two equations involving the constants a and b were obtained. These, in turn, were solved simultaneously and the constants evaluated. The resulting equations for cones and rods are given below.

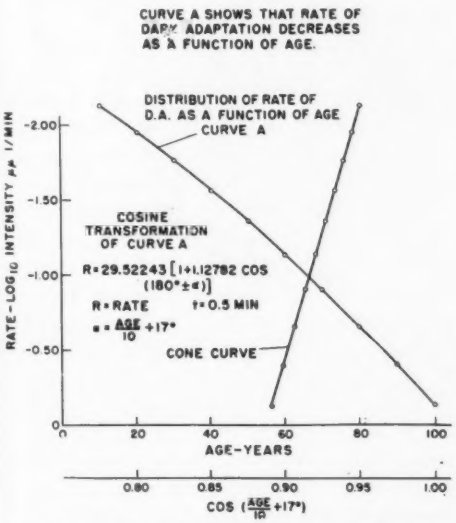

Fig. 14. Rate of dark adaptation is shown to be inversely related to age at time 30 sec (cones). See extrapolation to age 5 and 100 years. The cosine transformation results in a straight line.

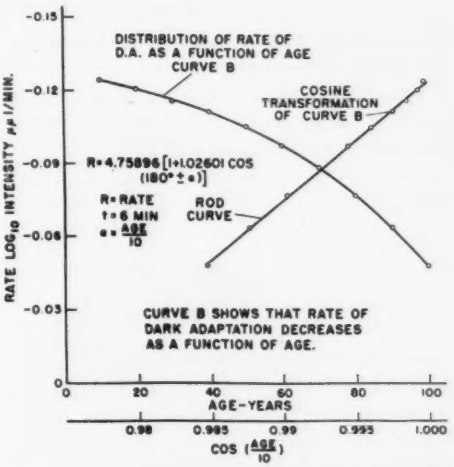

Fig. 15. Rate of dark adaptation is shown to be inversely related to age at time 6 min (rods). See extrapolation to age 5 and 100 years. The cosine transformation results in a straight line.

For cones at 30 sec and rods at 6 min:

$$\text{Cone rate} = 29.52743[1+1.12782\cos(180°+\phi_1)]$$

$$\text{Rod rate} = 4.75896[1+1.02601\cos(180°+\phi_2)]$$

Since ϕ_1 and ϕ_2 contain the age variable, rate will change as ϕ_1 and ϕ_2 change. It follows that dark adaptation is a lawful function of age. The full mathematical treatment of the data will be furnished the reader upon request.

REFERENCES

1. HAMMOND, C. E. and LEE, R. H., Criteria for Scoring Dark Adaptation Tests. Rept. No. 3 on Studies in Aviation Medicine Res. Section, July, 1941.

2. MCFARLAND, R. A. and FISHER, M. B., Alterations in Dark Adaptation as a Function of Age. *J. Gerontol.* **10** (4) 424, 1955.

Target Size and Visual Recognition[1]

WILLIAM C. STEEDMAN and CHARLES A. BAKER, Aerospace Medical Laboratory, Wright Air Development Division.

This study was conducted to determine the speed and accuracy of form recognition as a function of the size of target forms for various amounts of detail resolution. The stimulus forms were generated by filling in, on a statistical basis, some of the cells of a 90,000-cell matrix. The subjects were shown a "briefing target" and instructed to locate that target on a display containing numerous other forms. The significant finding is that both search time and errors remain invariant until the visual angle subtense of the targets falls below 12 min; at values below 12 min performance deteriorates. This effect is independent of the range of resolutions investigated. The implications of these findings to equipment design are discussed.

INTRODUCTION

This report is the second of a series emanating from a laboratory research program designed to investigate the recognition of targets such as those displayed on air-to-ground radar and infra-red sensing systems. The first report of this series was primarily concerned with target recognition as a function of resolution and the number of irrelevant targets displayed.[1] The object of the present study was to investigate the speed and accuracy of target recognition as a function of the displayed size and resolution of targets. The observer was briefed only on the form properties of a specific or briefing target to be located, while the location of the briefing target with respect to other displayed forms was unknown to the observer. The task of the subjects was similar in many respects to the task of an observer in an armed reconnaissance mission in which critical targets on the display must be recognized primarily by their inherent form or pattern properties.

The literature concerning visual acuity functions provides information relating to the visual angle subtense required of forms for accurate recognition. However, most of the investigations on these acuity functions are restricted to Landolt rings and parallel bar gratings. A review of this literature as it bears upon form recognition is contained elsewhere.[2] A general statement based on previous visual acuity studies would be that, under ideal conditions, the critical detail of a form which differentiates it from other forms must have a visual angle subtense of at least 1 min. Application of this finding is made difficult by the lack of a technique to describe, in a quantitative way, the nature of the critical detail of a form which can be used to differentiate it from a class of other forms. This is particularly true when dealing with forms of the nature of operational targets of military significance.

Clinical measures of visual acuity commonly use the absolute recognition of alpha-numeric symbols, e.g. the Snellen acuity chart. Individuals with normal vision are just able to recognize letters of the alphabet which subtend an overall visual angle of 5 min, but it should be remembered that symbols of the alphabet are overlearned and are readily discriminable one from the other. For the purposes of this investigation it was felt that the visual angle subtense of a form provided a reasonable means for studying speed and accuracy of form recognition. This study was designed to investigate the ability of subjects to recognize a particular target form occurring on a display containing many irrelevant forms, as a function of the visual angle subtense of the critical target form and of the amount of detail resolution.

STIMULUS MATERIALS

Target Form Generation

As in the previous study,[1] the stimulus materials were taken from a basic 90,000-cell (300×300) matrix, certain cells of which had been design-

[1] The authors are indebted to Dr. Clarke Crannell of Miami University, Oxford, Ohio, for providing supervision of the data collection and data analysis and to Dr. W. D. Chiles of the Aerospace Medical Laboratory for his editing and criticism of the draft manuscript.

ated as pattern (filled-in) cells as follows: each cell of the matrix was assigned two digits (from 0 through 9) at random; cells were then filled in as a result of either of two procedures. First, a cell would become a pattern cell and be filled in if both of the digits assigned to it were zeros. Second, a cell would be filled in if the second of the two digits were a 0, a 1 or a 2, and if, in addition, it touched the left, the right, or the bottom or the lower left or the lower right corner of a cell which had been previously filled in as a result of either the first *or* second procedure. The empirically determined probability that any given cell would become a pattern cell when these procedures were followed is 0.059. The resulting total 90,000-cell matrix which contains 557 discrete forms is illustrated in a previous issue of this journal.[1]

Problem Displays

The problem displays consisted of circular search areas, cut from the basic 90,000-cell matrix. Circular problem displays, 7.8 in. in diameter, were cut from copies of the basic matrix which were 12 in. on a side. Thus each problem display covered 33 per cent of the basic matrix. Since the basic matrix contained 557 discrete forms, each display contained about 184 forms. Five circular displays were selected so that, within the limits of the geometry of the situation, they covered the entire basic matrix. The five problem displays were then photographically reduced so as to result in a series of displays with diameters of 1.95, 3.90 and 5.85 in. respectively, plus the original 7.8-in. size. This yielded a total of 20 problem displays. The dimensions of the sides of the individual matrix

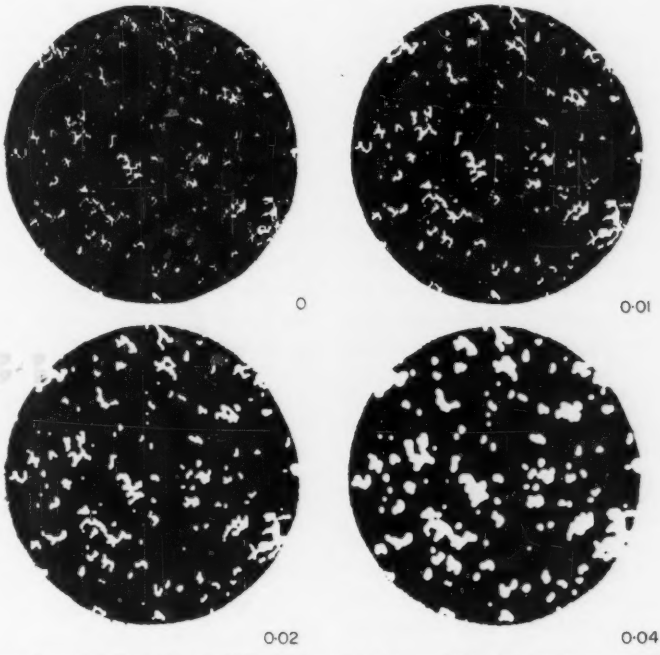

Fig. 1. The general appearance of the problem display matrix forms under the resolution conditions used in the experiment. The numbers refer to the blur disk diameters measured in inches. The matrix cell size depicted is approximately 0.02 in. In the study, the problem display size for this matrix cell size is actually an area 2½ times as large as the displays shown in the figure, and thus included approximately 2½ times as many forms.

cells were then 0.01, 0.02, 0.03 and 0.04 in., respectively, for the four problem display sizes.

Target Form Resolution

The resolution of the matrix display was controlled with the aid of a photographic contact printer by varying the exposure time, the space between the negative and the positive photographic plates, and the space between the negative plate and the extended light source. The loss in resolution for a given condition was measured in terms of the increase in the width of a single matrix cell as measured by a calibrated stage microscope. This measurement provides an approximation of what is usually called the "blur disk" diameter. The resultant values were checked against those obtained by measurements made with the Air Force Standard Resolution Pattern Negative Plate transformed into blur disk diameter values. These two independent calibration procedures resulted in nearly identical measurements of blur disk diameter. Four resolution conditions were investigated. For the largest problem display (cell dimension of 0.04 in.), the blur disk diameters were 0.00 ("perfect resolution"), 0.02, 0.04 and 0.08 in. For the smaller problem displays, these blur disk diameters were proportionately reduced. Note that, regardless of problem display size, the worst resolution condition has a blur disk diameter twice the dimension of an individual cell side. Fig. 1 illustrates the appearance of the four resolutions used in the study. The cells represented in the figure are 0.02 in. on a side. In the study, the problem display size for this matrix cell size actually included an area $2\frac{1}{2}$ times as large as the displays shown in Fig. 1, and thus included approximately $2\frac{1}{2}$ times as many forms.

Target Selection

A form was defined as either a single cell in isolation or a group of interconnected cells. From the total of 557 forms in the basic matrix, 24 forms were selected as targets. The targets were chosen to be representative of the total population and also with the following restrictions: (1) a form would not be selected as a target if, as a result of the "growth" encountered in the poorest resolution, it would touch another form, (2) a target could not appear more than once on a problem display, and (3) a target would not be chosen if it resulted in an uneven distribution of targets over the problem displays. Since the problem displays were rotated by 90° increments in successive presentations, each target appeared in each quadrant of the problem display. The 24 targets were cut from the basic matrix and mounted on a black background and placed on poster-board, then covered with transparent plastic. These single targets, called briefing targets, are shown in Fig. 2.

Fig. 2. The 24 targets used in the experiment. Target sizes range from three cells (No. 6) to 78 cells (No. 1).

EXPERIMENTAL DESIGN

A 4 × 4 factorial design was used. The independent variables were:

(1) Size of the matrix cells: the basic dimensions (for the highest resolution) of the matrix cells were 0.01, 0.02, 0.03 and 0.04 in.

(2) Resolution of the matrix cells; for the largest cells (0.04 in.), the blur disk diameters are 0.00, 0.02, 0.04 and 0.08 in. The blur disk diameters for the smaller cell-size conditions are proportionately reduced.

Four groups of 16 subjects each were used in the experiment, each group taking one-fourth of the total of 16 treatment combinations. Four subgroups of 4 subjects each had a different order of presentation of that group's 4 experimental conditions, as well as a different order of presentation of the 24 targets under each individual condition. This procedure provided counterbalancing of practice effects. Four different orientations (90° increments) of the problem displays were used to decrease the possibility of learning the actual target positions. In each case, however, the orientation of the briefing target was the same as that of the target as it appeared in the problem display. The criterion measures used were the mean and the median search times and the total number of errors, i.e. misidentifications. These criterion measures are based on the performance of the 16 subjects assigned to each experimental cell for each of the 24 targets.

PROCEDURE

The subjects were pretested on 16 targets (not used as briefing targets in the experiment proper) located on problem displays of the sizes used in the experiment. All the various resolutions were used. The pretest served three purposes: (1) to divide the subjects into 4 subject groups of nearly equal ability, (2) to familiarize subjects with the task, and (3) to screen out those subjects who found the task to be nearly impossible to perform. Of a total of 84 subjects pretested 20 were eliminated on the basis of poor performance.

One subject was tested at a time. The subject was seated facing a display board which was tilted at an angle of 45°. A viewing distance of 24 in. was maintained by a fixed chin rest. A diffuse illumination of 20 ft.-candles on this work area was provided by indirect lighting carefully positioned to minimize specular reflections. The experimenter placed the problem display on the display board. When the briefing target had been positioned immediately above the problem display, the subject was instructed to begin and a stop watch was started. The subject's task was to locate on the problem display the form which matched the briefing target and indicate his choice by marking the chosen form with a grease pencil. If the choice were correct, the time was recorded and a new problem display and briefing target were presented. If the subject's identification of the target were incorrect, he was so informed and instructed to resume searching for the correct target.

In a single session, which lasted approximately half an hour, the subject was given a 16-target warm-up trial followed by the 24 targets in the experiment. Each subject participated in 4 experimental sessions for a total of 96 identifications. A monetary incentive plan was used to maintain a high level of motivation among the subjects. Under this plan, a subject was rewarded 5, 10 or 15 cents, depending on his speed, immediately after making a target identification. However, the subject was also penalized from 5 to 25 cents for requiring too much time to select the target, and a 15-cent penalty was inflicted immediately for an incorrect target selection. Because of the large intersubject differences, sliding reward–penalty criteria were used (without the subject's knowledge) to balance the rates of pay so as to insure that the total rewards could at least equal the total penalties for any given subject. However, these criteria were chosen so that the faster subjects would receive a considerably larger reward than slower subjects over the course of the experiment. For a single session, a slow subject might earn no reward, whereas the faster subjects could earn a reward as much as two dollars. The reward–penalty totals were posted so that a subject could compare his performance with that of other subjects. This

incentive plan appeared to create a very high motivation in most subjects.

RESULTS

Matrix Cell Size and Resolution

The dependent variables are search time and errors. The mean and median search times were determined for each target in each of the 16 experimental conditions. Each of these basic mean and median scores is based on 16 observations, one by each of 16 subjects. All summary data are based on the means of these mean and median scores. The effects of matrix cell size are shown in Fig. 3. Both search time and errors are invariant for the three largest cell sizes but increase considerably for the smallest cell size. The median search times from the highest to poorest resolution are 11.8, 10.3, 10.3 and 11.9 sec, respectively. Thus it is concluded that resolution had little or no effect upon search time. This latter finding agrees with the first study of the series, which used the same range of resolutions.[1] There is also no apparent interaction of resolution with matrix cell size. In addition, errors do not vary systematically with resolution. Therefore, in the subsequent analysis the resolution variable will be ignored.

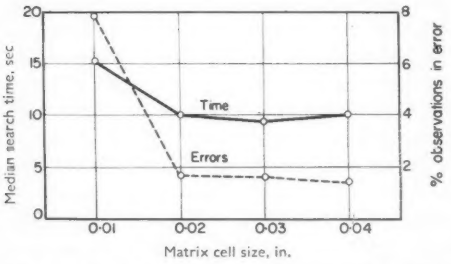

Fig. 3. Median search time and percentage of observations in error as a function of matrix cell size. Each point on the graph was derived from 1536 observations.

Of significant interest to this investigation is the rise in search time and errors for the smallest cell-size condition. To obtain a better understanding of this effect a more detailed analysis was carried out. First, the performance data for each target for each cell-size condition were compiled. (See Table I.) Examination of these data reveal that not all of the 24 targets produced an increase in errors and search time for the smallest cell-size. The affected targets are the smallest ones, i.e. contain the fewest number of matrix cells, whereas the performance on the larger targets was relatively unaffected by the matrix cell size. Second, the maximum dimension of each target for each matrix cell size was computed. Since viewing distance was held constant (24 in.), these maximum dimensions could be converted into minutes of visual angle subtense which would provide a more convenient metric for purposes of generalization. Since targets varied in relative difficulty and since this was independent of the cell size, it was necessary to use a relative performance measure. To obtain this measure, the median search time for each target for each matrix cell size condition was compared to the overall median search time of that target, and the difference then expressed as a percentage of increase or decrease in search time. Since 24 targets were used in each of 4 matrix cell sizes, 96 scores were obtained. The 96 target cell-size conditions were then ordered by visual angle subtense. These 96 conditions were arranged in groups of eight, starting with the smallest visual angle, i.e. the first group contained the 8 target conditions with the smallest visual angle subtenses; the second group, the next eight, and so forth. Each set of 8 visual angle subtenses was then averaged to provide a size measure to represent each of the 12 groups thus formed. The median search times associated with each of these 12 groups were plotted against the corresponding visual angles and are shown, together with the error data, in Fig. 4. It is evident that when the maximum dimension of the target is less than 12 min of visual angle, there is a precipitous rise in both search time and errors. It should be noted that 63 per cent of the total number of errors (122 out of 193) occurred in the smallest display condition, which represented only 25 per cent of the total observations. Further, those targets subtending

TABLE I

Performance Measures for the 24 Targets Studied. Each Error Entry is Derived from 256 Observations. Each Median Time is Derived from 64 Observations

Target number	*Visual angle**	*Number of errors*	*Median search times for the various matrix cell sizes*				
			0.01	*0.02*	*0.03*	*0.04*	*X*
1	31.5	4	3.1	2.9	2.4	2.2	2.7
2	22.3	1	4.2	3.5	3.9	4.6	4.0
3	14.0	0	8.2	6.7	7.0	5.8	6.9
4	11.6	0	4.8	4.7	2.9	2.4	3.7
5	8.2	21	25.1	15.2	16.1	14.2	17.7
6	6.0	5	8.9	4.3	4.5	7.7	6.4
7	21.7	0	3.9	4.1	4.6	4.6	4.3
8	6.4	12	24.7	11.7	11.8	14.7	15.7
9	9.2	2	7.3	9.5	8.3	12.2	9.3
10	28.6	0	4.0	3.0	2.6	2.6	3.1
11	7.7	25	51.7	19.9	16.9	20.0	27.1
12	7.2	1	14.2	7.6	4.9	6.6	8.3
13	9.7	1	7.2	7.9	6.0	7.5	7.2
14	12.2	2	10.3	5.9	6.0	4.9	6.8
15	7.2	8	16.7	11.5	12.0	12.7	13.2
16	6.6	12	20.6	10.0	13.6	13.0	14.3
17	7.7	10	18.8	12.5	11.1	9.6	13.0
18	9.0	11	12.5	13.8	12.7	11.8	12.7
19	8.3	19	41.6	28.6	24.5	26.7	30.4
20	11.2	23	16.7	13.5	11.0	13.7	13.7
21	6.6	6	21.3	10.1	9.7	12.0	13.3
22	8.2	13	8.7	9.0	5.9	5.2	7.2
23	9.7	4	14.0	12.5	12.7	9.8	12.3
24	6.2	13	19.5	9.8	12.8	14.0	14.0

* These visual angles in minutes of arc are based on the maximum dimension of each target for the 0.01-in. cell size, high resolution condition. These values should be proportionally increased for the larger cell size conditions. (It should be noted that as the blur disk diameters are increased, the visual angle sizes are also increased slightly. The data analyses in this report ignored this slight increase in visual angle resulting from the blur disk diameter.)

less than 12 min (18 of 24 targets) contributed 98 per cent of the errors in that condition. It may be presumed that 12 min of visual angle subtense is the lower limit that is acceptable for target recognition for the type of forms used in this study.

Target Orientation

It was pointed out earlier in this report that the briefing targets and problem displays were rotated by 90° steps through 360°. In each case, however, the orientations of the briefing target and that target on the problem display were identical. As could be expected, search time and errors did not vary for the various orientations used.

It was of interest to determine how performance would be affected if the orientation of the critical target on the problem display were unrelated to the orientation of the briefing target presented to the subject. Therefore, 8 additional subjects with pretest scores nearly

equal to those of the original 64 subjects were selected. These 8 subjects were run in an experiment in which the orientation of the target on the problem display was unrelated to the orientation of the briefing target. The appearance of the targets by quadrants was balanced,

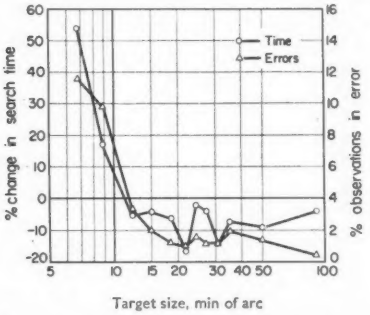

Fig. 4. Relative increase or decrease in median search time and frequency of errors as a function of the visual angle subtense of the targets. Each point was derived from 512 observations.

as in the experiment proper, by rotation in 90° steps through 360°. All four resolutions were used, but only the matrix cell size of 0.03 in. was used. Using the scores obtained from the appropriate conditions of the study proper, comparison revealed that search time increased 101 per cent when the subjects did not know the orientation of the target in the problem display. Frequency of errors increased from 1.6 per cent to 9.1 per cent. These increases were about constant for each of the four resolutions investigated. Since the subject samples and general procedure of the two studies were similar, it is safe to conclude that the search times reported in the experiment proper would be approximately doubled and errors increased by a factor of about five if the orientation of the targets on the problem display were unknown.

Target Analysis

The search times for the different targets were found to exhibit a large amount of variability. The easiest target (No. 1) had a median search time of 2.6 sec; the most difficult target (No. 19) had a median search time of 30.3 sec. This relative target difficulty was also revealed in the occurrence of errors. For example, no errors of identification occurred in 256 observations for each of 4 targets, as compared with 25 errors in 256 observations of target No. 11. A correlation of 0.81 between time and errors was found, which indicates that those targets which required the longer search times were also misidentified more frequently.

Practice

Since each of the 64 subjects had 4 of the 16 experimental conditions in a counterbalanced design, it was possible to analyze the effects of practice over the four trial days. Fig. 5 indicates that practice does have an effect on search time. In the previous study the subjects had 16 trials in which the decrease in search time was approximately linear until after the eighth trial session.[1] Thereafter, the decrease in search time with additional practice began to level off. Thus, the search times reported in the present investigation are considerably larger than would be expected if the subjects had been given more training at this particular task.

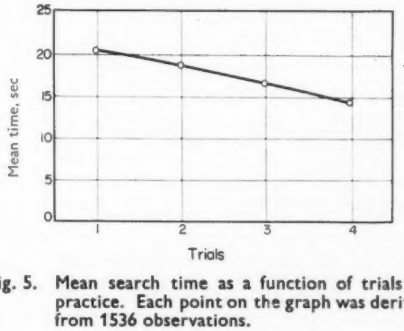

Fig. 5. Mean search time as a function of trials or practice. Each point on the graph was derived from 1536 observations.

DISCUSSION

The significant finding was that both criterion measures, namely, search time and errors, remain relatively invariant until the visual angle subtense of the maximum dimension of the targets falls below 12 min. Assuming a 12-in.

viewing distance, a target must have a minimum size of 0.042 in. as displayed in order to expect relatively accurate and rapid recognition. Thus, given the scale factor of an analogue display, such as a radar navigation display, one can compute the actual dimension of ground targets required for acceptable visual recognition on the analogue display. Fig. 6 gives a plot of display size against target size for 8

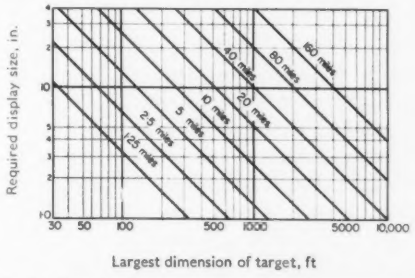

Fig. 6. Required display size plotted against target size for various ground ranges displayed to the observer. These functions assume a viewing distance of 12 in. to the display and a visual angle subtense of 12 min for the largest dimension of the target to be recognized. See text for fuller explanation.

values of ground range. Individual values may be computed by the following formula:

$$\text{Display size in inches} = \frac{\text{ground range}}{\text{target size}} \times 0.042$$

This plot assumes that the target must subtend a visual angle of 12 min in order to be recognized, and that the viewing distance to the display is 12 in. Thus, if the smallest target one needs to recognize is 1000 ft in its greatest dimension, and the system displays an analogue of a strip of ground 40 miles wide, the display must have one dimension of not less than 10.2 in. On the other hand if a critical target has a length of 50 ft, and the display size is fixed at 10 in., the ground range displayed cannot exceed 2.1 miles.

The application of the findings contained in this report to operational radar or infra-red displays is restricted by several considerations. In this study, although the subjects did not know where the critical target was with respect to the other displayed forms, the briefing target shown them was an exact replica in size and shape of the target as it appeared on the problem display. Thus, in this respect, target intelligence was perfect. Also, the subjects in the experiment knew that the critical target was *always* present on the problem display. If the experiment had been conducted so that the presence of a target was uncertain, as is the case in military operations, a source of error not possible in the experiment as conducted could have occurred, namely, the subjects could have erred by reporting no target present on the display when in fact there was a target present. Although the actual target forms used were statistically generated, it is felt that this is not a serious limitation in generalizing the results of the study to operational targets.

It must be remembered that the value of 12 min of arc for accurate absolute target recognition assumes nearly perfect target intelligence with respect to form characteristics, and also high contrast and ideal viewing conditions. An operational environment almost certainly will provide less than optimal conditions, such as, less than perfect intelligence relating to the form characteristics of the displayed target, greater similarity between non-target forms and target forms, vibration and buffeting and operator fatigue. Based on scattered elements of information in addition to the data from this experiment, the 12 min of visual angle may be unrealistically small when the operational environment is considered. Thus, where practical, the minimum visual angle should probably be approximately 20 min, rather than the 12 min found in this investigation.

REFERENCES

1. BAKER, C. A., MORRIS, D. F. and STEEDMAN, W. C., Target recognition on complex displays. *Human Factors*, **2**, (2), 51, 1960.

2. HAKE, H. W., Contributions of Psychology to the Study of Pattern Vision. Technical Report 57-621, Wright Air Development Division, 1957.

A Laboratory Evaluation of the Effects of Electronic Countermeasures on System Performance[1, 2]

WILLIAM D. HITT and HORACE W. RAY, Battelle Memorial Institute, Columbus, Ohio.

The present research program was designed to make optimum use of selected laboratory and analytical methods for purposes of simulating a relatively complex man-machine system. Results of this research program indicate that the laboratory approach, when applied within a systems framework, offers an efficient and meaningful method for the evaluation of ECM effectiveness.

INTRODUCTION

If our strategic bombers were to fly over enemy territory, the enemy would attempt to detect and locate them by means of their ground radar system. Well-defined "blips" appearing on the enemy radar scopes would make these bombers extremely vulnerable targets for interception. To distort or obscure the enemy's radar signals, our modern bombers would carry special electronic gear called ECM (for electronic countermeasures). The sole purpose of the ECM would be to deny the enemy the information needed to intercept and destroy our bombers.

Various types of electronic countermeasures have been designed for use against the many types of radar equipment which a modern defense system probably would include. Electronic countermeasures have also been designed to combat the attempts of the ground radar system to resist electronic jamming of their radars by using various electronic counter-countermeasure techniques. Thus, to plan a successful bomber penetration of enemy territory, it is necessary to select the specific electronic countermeasures (ECM) most likely to degrade the performance of the ground radar system, which may or may not be making use of electronic counter-countermeasures (ECCM).

The choice of electronic countermeasures is based on so many complex and interrelated factors that such decisions have become the subject of an extensive body of scientific research. In research on the relative effectiveness of various types of ECM, two of the more difficult decisions to make concern the selection of a criterion to use as a measure of ECM effectiveness and the selection of a simulation method.

The most frequently used criteria of ECM effectiveness represent measures of radar operator performance: an ECM technique is considered effective if it degrades such scores as detection time, blip-to-scan ratio, azimuth error, range error or heading error. A major deficiency of such immediate criteria, however, is that they provide practically no information concerning the extent to which the crucial part of the air defense system performance is impaired by the ECM; i.e. they do not predict bomber survival.

An equally important problem in ECM evaluation is the method used to simulate the system. An extremely realistic method might include a real ground radar system operating against real bombers, or a method extremely low on the "reality continuum" might be a completely digital simulation. Operational field tests are indeed realistic, but the expense and lack of experimental controls prohibit extensive use of this method. A completely digital approach could be well controlled, but the ground operator cannot yet be simulated on a digital computer. Thus, for purposes of this study, the selected method falls between these two extremes on the degree-of-reality continuum.

1 This research was conducted by Battelle Memorial Institute, Columbus, Ohio, under subcontract to General Mills, Incorporated, Minneapolis, Minnesota. The Prime Contract was U.S. Air Force Contract AF 33(616)-3739.

2 The authors of this paper were members of a four-man Battelle research team. A. A. B. Pritsker developed the intercept model, and B. B. Gordon was responsible for the development of the simulation equipment.

The major objective of the present research program was to develop a simulation facility for use in evaluating ECM effectiveness. This paper has been written primarily to demonstrate the simulation of a relatively complex man–machine system by laboratory and analytical techniques. To illustrate how such a method can be applied, one of the studies conducted at Battelle to evaluate ECM effectiveness will be described.

the simulated target-return signal (bomber blip) and the ECM generated by the bomber as altered by ECCM techniques used by the ground radar.

Human subjects, acting as radar operators, tracked the bomber blip and reported the observed bomber co-ordinates to a key-punch operator. The true bomber co-ordinates and time were automatically punched into the same card. By use of these observed and true co-

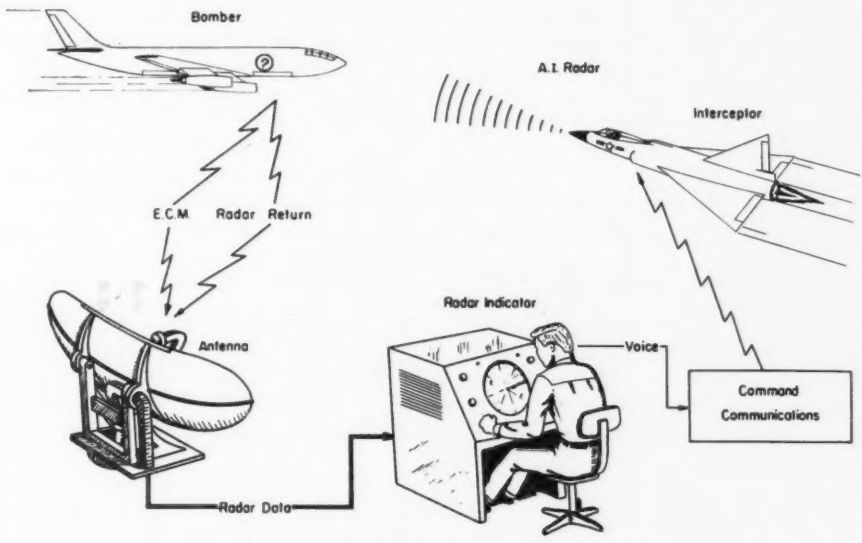

Fig. 1. Pictorial sketch of the simulated system.

METHOD

The Simulated System

Fig. 1 is a pictorial sketch of the air defense subsystem simulated in the present study. This subsystem consists of three major components:

(1) An attacking bomber carrying ECM,
(2) A ground radar system, and
(3) A ground-controlled interceptor.

The simulated radar system with its resultant radarscope display was created by use of an analog computer, an ECM and radar simulator and related equipment. The display presented ordinates, an air battle between an attacking bomber and a ground-controlled interceptor was simulated on a digital computer intercept model.

For purposes of this research program, ECM composition represents the independent variable in the simulated system. The measure of effectiveness of the ECM, or the criterion, is the probability-of-kill score (PK) provided by the intercept model. Viewing the situation from the standpoint of the attacking bomber, the effectiveness of the ECM realistically determines its "probability of survival", which is simply 1.00 – PK.

Radar Simulation

The following steps were taken to achieve realism in the simulation. Target–signal strength and size were varied with changes in target range and target echo area. Antenna beam width was taken into account in setting the r-f level as it actually would be applied in a real receiver.

A general–purpose analog computer produced the required outputs of the target generator. In addition, certain functions of the sweep and scan driver, the recording system

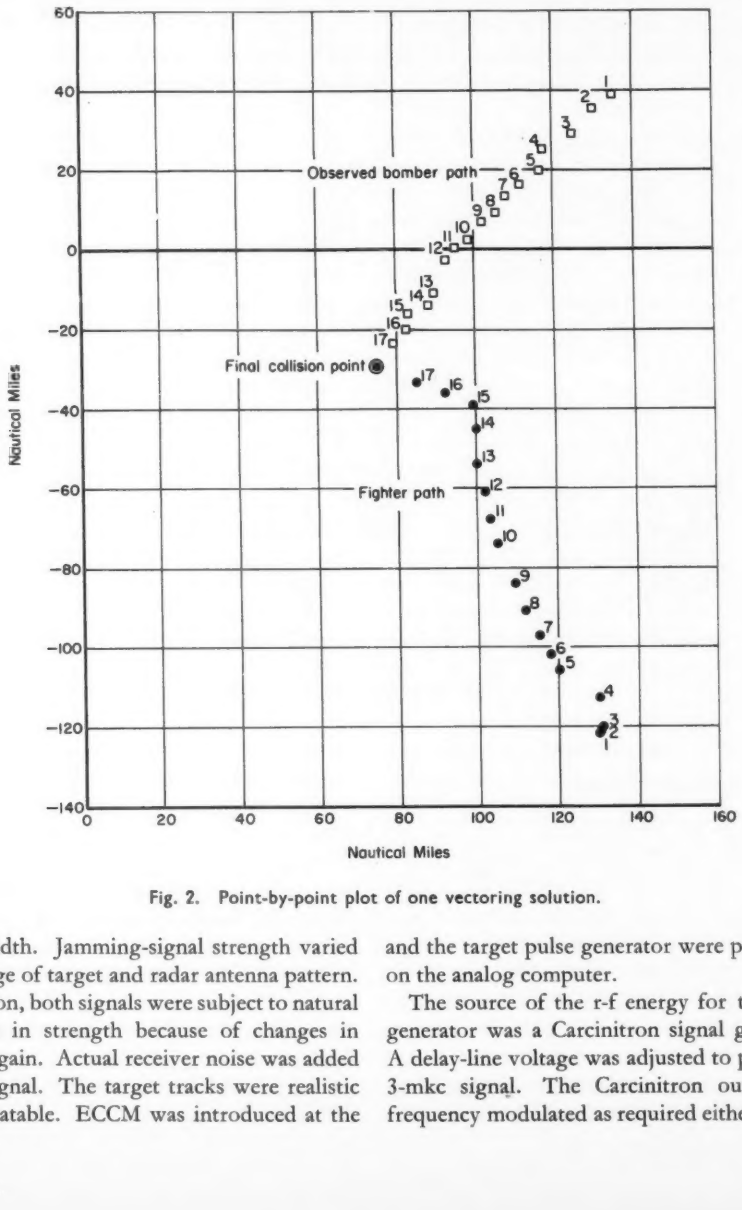

Fig. 2. Point-by-point plot of one vectoring solution.

target width. Jamming-signal strength varied with range of target and radar antenna pattern. In addition, both signals were subject to natural variation in strength because of changes in receiver gain. Actual receiver noise was added to the signal. The target tracks were realistic and repeatable. ECCM was introduced at the and the target pulse generator were performed on the analog computer.

The source of the r-f energy for the ECM generator was a Carcinitron signal generator. A delay-line voltage was adjusted to produce a 3-mkc signal. The Carcinitron output was frequency modulated as required either by pure

noise or by sweep plus noise. This output was then amplitude modulated in a traveling-wave-tube amplifier by a signal proportional to the radar antenna-pattern gain.

The ECM signal and the target signal, after being properly attenuated, were mixed in a hybrid junction (Magic T). The signal from the junction was mixed with the local-oscillator signal in a balanced mixer. Matched crystal detectors fed the inputs of a balanced-input preamplifier. The preamplifier output fed the receiver, which consisted of a broadband i-f amplifier, a narrow-band filter, a postamplifier and a detector. When desired for the experiment, the ECCM was inserted between the i-f amplifier and the filter.

The Intercept Model

A second major step in the system simulation was the development of a digital model to simulate an air duel between one bomber and one interceptor.[1] The model covered the time period from interceptor take-off to intercept armament release. The criteria for the interceptor path were that the interceptor should fly an initial approach leg, make a turn consistent with its aerodynamic limitations and turn into a final closing straight leg at target altitude and at interceptor speed.

The factors considered in the intercept model were the range between the interceptor and the bomber, the aspect angle between the interceptor and the bomber, the AI radar system parameters, the AI radar echo area of the bomber and the visibility factor. If these conditions were optimum for interceptor performance, then a high probability-of-kill score was obtained, whereas if these conditions were less than optimum, a somewhat lower probability-of-kill score was obtained.

Fig. 2 portrays a point-by-point plot of the time history of one vectoring solution.

AN ILLUSTRATIVE EXPERIMENT

Experimental Design

Two types of ECM and one type of ECCM were evaluated. Selection of the specific techniques studied was based on the requirement that the results of this study be comparable with the results obtained previously by use of a different approach. The selected ECM techniques were FM by noise and swept spot. For purposes of security, the particular type of ECCM selected for this study will not be specified.

It was originally planned to vary each type of ECM at four different power levels, operating against simulated systems that made use of the ECCM and against those that did not. However, pilot studies showed that the ECCM practically eliminated the effects of swept-spot jamming even at the higher ECM power levels. Hence, it was apparent that it would not be necessary to evaluate the swept-spot jamming—ECCM combinations for more than one power level.

As a control condition, a no ECM—no ECCM combination was included in the experimental design. Under this experimental condition, only the target-return signal was presented on the radarscope.

To achieve a completely balanced Graeco-Latin square design, an odd number of experimental conditions was desired; consequently, an additional power level was included for the swept-spot jamming—no ECCM combination.

The ECM power levels employed in the study varied from -5 dB to -13 dB. (These numbers refer to an attenuator setting in the simulation and are meant only to give comparative power information.) As mentioned previously, jamming-signal strength varied with range of target and radar antenna pattern. Therefore, the reported power levels represent the levels for only the initial phase of an experimental trial.

The 15 experimental conditions are presented in Table I.

Individual differences in tracking ability, learning effects and bomber tracks, coupled with the major independent variable of 15 experimental conditions, led to a 4-variable design. By establishing the same number of "levels" for each of the four variables (i.e. 15), it was possible to produce a Graeco-Latin square design.

To control for individual differences in tracking ability, each of the 15 subjects tracked the bomber blip for one trial under each of the 15 experimental conditions. To control for learning effects, each subject's initial trial was given under a different experimental condition. Subject A, for example, started with Condition 1, Subject B with Condition 2, and so on. The mean subject performance for all experimental conditions therefore represents comparable levels of tracking proficiency.

The third extraneous variable, bomber track, was controlled by presenting each of 15 different bomber tracks under each of the 15 experimental conditions. These 15 bomber tracks consisted of 12 curved tracks and 3 straight-line tracks. Examples of 2 of the bomber tracks are shown in Fig. 3. The legend indicates those characteristics common to all curved bomber tracks. As in the case of curved tracks, straight-line tracks were maintained at bomber speeds of 500 mph.

Training of Subjects

Fifteen male Battelle technicians were selected as subjects for the experiment. A training program was established for the purpose of training the subjects to perform as radar operators. The program consisted of 2 types of training: (1) a general orientation on radar, ECM and tracking, and (2) an "on-the-job" training session on tracking procedures.

The general orientation meeting consisted of three major parts:

(1) Radarscope interpretation,
(2) Practice in estimating azimuth and range by completing a paper-and-pencil test, and
(3) Presentation of slides depicting jamming patterns.

On-the-job training consisted of the following:

(1) Explanation of the radar display,
(2) Explanation of procedure to be followed by subject,
(3) Demonstration of procedure by experimenter, and
(4) Completion of practice trials by subjects (with and without jamming).

Experimental Procedure

Each subject was dark adapted for 10 min prior to taking his position at the radarscope. His first task following the starting signal was to detect the bomber blip. He would then read azimuth and range at approximately 30-sec intervals—if he could see the blip. (The subjects were instructed not to give the co-ordinates unless they were "fairly certain" they actually saw the blip.)

The observed co-ordinates were conveyed by intercom to a key-punch operator, who recorded them. The experimenters also recorded the readings by hand so that the punched cards could be checked for correctness. At the same time, the true co-ordinates, along with time and a code identifying the run, were punched automatically into the same card. Three 15-min trials were made during each experimental session.

This procedure was followed until all 15

TABLE I

The 15 Experimental Conditions

ECM power (−dB)	*Condition no.*			
	No ECCM		*ECCM*	
	Swept spot	*Noise*	*Swept spot*	*Noise*
5	1	7	6	11
7	2	8		12
9	3	9		13
11	4	10		14
13	5			

	No ECCM
No ECM	15

subjects had tracked under all 15 experimental conditions, resulting in a total of 225 trials. For each trial there was a set of IBM tabulating cards, which contained the data used to obtain the measure of ECM effectiveness—probability of kill.

Consider, for example, the mean PK of 0.06 obtained for the swept-spot jamming, no ECCM, −5 dB display. This value may be meaningfully compared with the value of 0.75 obtained for the swept-spot jamming, ECCM, −5 dB combination. The difference between

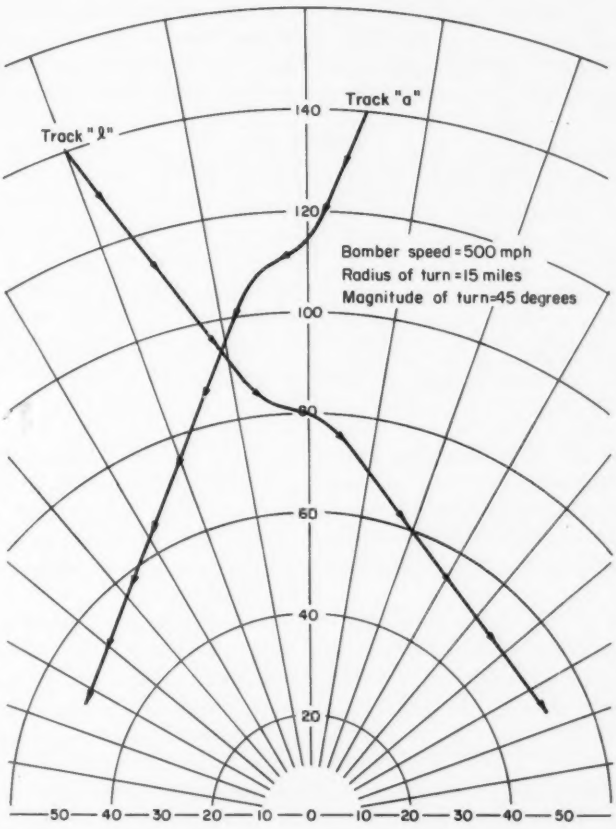

Fig. 3. Examples of two bomber tracks used in the experiment.

Results

For 87 of the 225 experimental trials, the subjects either did not see the bomber blip during the entire 15-min trial or else did not see it enough times to establish a track. These trials were automatically assigned a PK of 0.00. The mean PK's for the 15 experimental conditions are shown in Table II.

these two figures demonstrates rather vividly that, while swept-spot jamming at this power level is extremely effective against a normal receiver, it is apparently valueless if the ground radar is using this particular ECCM technique. This conclusion is further confirmed by comparing this mean PK of 0.75 with that of 0.80 obtained under conditions of a clear scope.

Figs. 4 and 5 represent graphically the results reported in Table II. It can be seen in Fig. 4 that, as ECM power level increases, PK decreases rather rapidly for both noise and swept-spot jamming under the condition of no ECCM. Swept-spot jamming appears to be slightly more effective than noise jamming, but the magnitude of this difference is not statistically significant.

Fig. 5 demonstrates that increasing the power level for noise jamming causes a rapid decrease in PK, even if the ECCM is being used by the ground radar. As mentioned previously, however, swept-spot jamming apparently is almost completely ineffective against this particular ECCM technique, even at a high ECM power level.

It is important to note that in most cases a change in ECM power of only 6 dB produces an extreme change in PK. These power levels are difficult to establish in reality, however, since they depend on many factors in the system, such as radar power level, radar antenna gain and desired jamming range.

CONCLUSIONS

It has been implied throughout this paper that probability of kill represents an "ultimate" criterion of system performance. PK, as used here, of course, is a simulated ultimate criterion of system performance. To obtain an actual ultimate criterion, it would be necessary to conduct operational field tests in which interceptors launched their rockets against attacking bombers. Obviously, such an approach would be impossible. Even with field tests that stop short of rocket launching, there are major problems with which to contend. First, the monetary cost of such tests is tremendous. Second, it is almost impossible to adequately control the pertinent variables during operational field tests.

It appears, then, that the laboratory approach is the most efficient method for evaluating ECM effectiveness. Even though the simulation of system components in the laboratory does not provide exactly the same system that exists in the field, it is possible to at least approximate the conditions that exist in an air defense

TABLE II

Mean PK for Each of the 15 Experimental Conditions

ECM Power (−dB)	Average PK: No ECCM, Swept spot	Average PK: No ECCM, Noise	Average PK: ECCM, Swept spot	Average PK: ECCM, Noise
5	0.06	0.09	0.75	0.00
7	0.12	0.24		0.00
9	0.26	0.34		0.29
11	0.53	0.61		0.69
13	0.45			

	No ECCM
No ECM	0.80

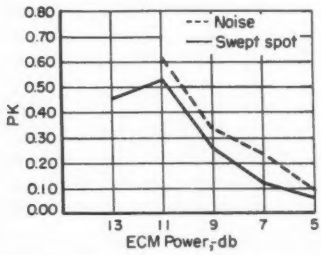

Fig. 4. Effect of ECM power on PK for both noise and swept-spot jamming—against no ECCM.

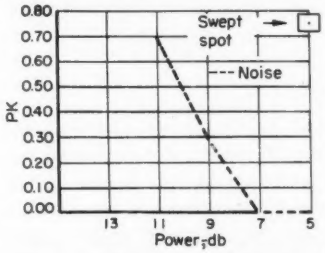

Fig. 5. Effect of ECM power on PK for both noise and swept-spot jamming—against ECCM.

environment. If future ECM simulation equipment can be sufficiently refined so that the settings can be converted to the settings on operational equipment, then the possibilities for evaluating ECM effectiveness under laboratory conditions appear to be unlimited.

One of the most important conclusions to be derived from the present research program is that it seems imperative that the evaluation of ECM effectiveness be done within a systems framework. If this research program had been designed to ascertain the effects of ECM on radar-operator performance, several measures of operator performance would have been available. In turn, the selected ECM displays could have been rank ordered on a degree-of-effectiveness continuum, according to the mean operator performance scores achieved under the various displays. Based upon previous work at Battelle, it is certain that the results obtained from such an approach would have been misleading. (Mean range errors and mean azimuth errors made by the operator, for example, appear to be unrelated to PK.)

It may be found from future research, however, that a more immediate criterion, such as detection time or blip-to-scan ratio, is just as valid as PK as a measure of ECM effectiveness. Nevertheless, such a finding will be possible only if PK, the ultimate measure of system performance, is included as a part of the research program. Moreover, it would seem that if a measure such as detection time or blip-to-scan ratio were to be used as a measure of ECM effectiveness, then the resulting scores should, nevertheless, be interpreted in terms of what effect the ECM has on *system performance*.

REFERENCE

1. Pritsker, A. A. B., Van Buskirk, R. C. and Wetherbee, J. K., Simulation to Obtain a Systems Measure of an Air Duel Environment. *IRE Transactions on Electronic Computers*, EC-8, p. 55, 1959.

A System Measure of Driver Performance[1]

DAVID B. LEARNER[2], Human Factors Research Group, Research Laboratories, General Motors Corporation.

A general review of measures of driving performance introduces a discussion of criteria requirements. Specific suggestions for criteria development related to psychomotor, intellectual and perceptual processes of the driver are included. In addition, a general measure for determining system performance that has application for all vehicular man–machine systems is described.

During the last fifty years there has been increasing concern on the part of various organizations with the problem of measuring driver performance. Law enforcement agencies are concerned because of their licensing function. Educational institutions are interested as a result of driver training courses. Manufacturers are concerned through their desire to develop a better product. Although the objectives of these organizations are quite different they are related by the common search for some objective, reproducible criterion of evaluating the performance of the vehicle–driver system.

This search has led to a number of approaches that have proven to be of value in specific situations. Broadly, these approaches are concerned with (a) driver selection, (b) driver training, and (c) safety. In terms of driver selection the candidate is evaluated on his knowledge of the rules of the road and his ability to negotiate a specified course to the satisfaction of the examiner. Driver training courses are evaluated in much the same way, in fact, quite often the ultimate criterion for driver training is the driver selection examination administered by state licensing bureaus.

Safety investigators have become more objective in their measurement techniques. Specific methods for evaluating reaction time and other complex processes have been utilized with some success. In a related field, driver behavior researchers have used measures of maneuver reproducibility and velocity control variability with good results. Of course there are probably some notable exceptions to the measurement of performance in each of these instances but in general they are concerned with directly measuring the driver's behavior in some specific circumstances and inferring how well he controls his vehicle.

With the development of the system concept it is appropriate to reflect on this problem in order to determine levels of measurement that will be most meaningful. Since man is highly adaptable and handling an automobile (in the broad scheme of things) is a rather simple task there seems to be a clear requirement that any criterion measure be sensitive to small variations in driver behavior. Furthermore it is fundamental that the performance measure include the performance of the driver in the vehicle, not merely the performance of the driver and the vehicle. This subtle difference suggests the necessity for measuring the output of the entire driver–vehicle system in such a way that the total system variation may be attributed to either the man component or the machine component or some interaction of the two.

This system measurement approach does not preclude the measurement of human subsystem responses. When the driver is viewed as a component, conventional human response measurements may provide even more meaningful information about the interface between man and machine. For example, it would be valuable to know the extent of human response that may be due to the system dynamics and the extent of human response that may be injected as noise. The conventional response data may be even more valuable when related to a higher level system performance criterion.

[1] A reproduction of a talk given before the Human Engineering Concepts and Theory Short Course at the University of Michigan, Ann Arbor, Michigan, August 10–21, 1959.

[2] Author is now Assistant Director of Research, Batten, Barton, Durstine & Osborn, Inc., New York.

Some of the requirements of driver performance measurement have been described. Beyond this there is a clear requirement for a measure of the output or response of the vehicle as a result of driver input. The driver performance requirements may be classified into three categories (a) psychomotor, (b) intellectual, (c) perceptual. Furthermore there is a clear requirement for measuring driver preference. It is not sufficient to assume that an operator prefers that system (whether it be a hammer or a data processing system) with which he performs best. Direct measurement of his preferences are necessary. This is essentially a process of quantifying qualitative variables. It is not inappropriate to point out that in addition to all other requirements whatever measures are used be reproducible and valid otherwise they are virtually useless.

In order to determine meaningful system response it is valuable to review the vehicle dynamics. The automobile may be described on a three-axis reference system. The fore and aft dimension is termed performance, the vertical dimension is ride and the lateral control dimension is handling. In this case we are concerned exclusively with the lateral response of the system. This lateral response is essentially any change in path of the vehicle. The automobile motions associated with lateral response are yaw, roll and sideslip. Yaw is the angular velocity about the vertical axis of the car. Roll is the angular velocity about the longitudinal axis of the car. Sideslip relates to the side velocity as defined by the sideslip angle (the angle whose tangent is sideslip velocity divided by forward velocity). Any lateral response of the vehicle will result in some change in these motions. A study of the response equations shows that it is possible to find modes of motion that are directional and rolling in nature. Since directional motion consists of yaw, and sideslip and system performance is related to directional change, it seems straightforward to view yaw (or sideslip, since there is a strong coupling between the two) as a measure of system performance in terms of position or path following. This may be viewed in another way. Whenever the vehicle changes position there is rotation about its vertical axis. This angular rotation has a velocity which depends to a great extent upon the vehicle's forward velocity and is called yaw rate. As a result the position of the vehicle may be derived by measuring yaw rate.

Furthermore the inherent measurement simplicity of this approach lends it great attractiveness. In summary, if vehicle yaw rate is measured at a constant forward velocity, the vehicle position (or system output) may be derived by:

$$P = \iint \dot{r}\, dt\, dt \qquad (1)$$

$\dot{r}$ = yaw rate.

Now that system output is directly available (from a yaw rate gyro) knowing the input or forcing function would allow computation of the system error or performance. In laboratory tracking situations this simple model describes the system error:

$$P_A - P_D = E \qquad (2)$$

where P_A = actual position

P_D = desired position

E = error.

Equation (1) has established that actual position may be derived from yaw rate. When desired position (or input) is specified as being zero displacement, from some course, then, by subtraction, the error is equal to the actual position. As a result, by measuring yaw rate, system error can also be measured.

In order to implement this procedure it is necessary to specify to the driver the task of driving straight down the road (zero displacement). This is in fact so simple that it is no task at all. In order to increase task difficulty so that system design differences might be detected the driver must perform the specified task under disturbing conditions. This disturbance is provided by a very low frequency complex wave that positions the front wheels through a servo and has much the same effect as road potholes or a brisk cross wind. Now the human

operator has a difficult task, attempting to drive straight in the face of disturbance. As a result, his desired position is zero displacement (zero yaw) and the extent of accumulated yaw rate (mean position error or third integral of yaw rate) is the actual error and is a figure of merit of system performance.

The extensive presentation of this method of measuring system performance is warranted from two different aspects. This approach achieves significant flexibility through its self-contained signal generating and error recording equipment. Any vehicle may be compared with any other vehicle over any stretch of road, providing, of course, that the comparisons are based upon data gathered over the same stretch of road. Secondly, the task is directly analogous to driving over a badly maintained road surface or while subjected to a variable cross-wind. Finally the simplicity of implementation is a benefit not usually resulting from this kind of measurement.

Thus far primary concern has been with developing the man–machine system criterion. Providing that this requirement has been satisfied, the problem of measuring the psychomotor, intellectual and perceptual performance of the operator still prevails. With regard to psychomotor performance it seems that frequency and amplitude measures of steering wheel movements could detect different performance demands upon the operator. Recent research [1] indicates that both of these dimensions play significant roles in the operators' motor response and that in order to combine them some measure such as total operator power dissipated might be appropriate.

Intellectual performance of drivers has historically been measured by intelligence type tests.[2] In this context, however, intellectual activities are primarily those complex processes that lead to decisions. As such, decision making becomes a rather vital topic for driver research. Relationships between individual probability preferences and risk taking have apparent application in many common highway situations. The utility of the goal that motivates a trip may tell us about the probability of making decisions that lead to accident fraught situations. The behavioral tendency to overestimate the occurrence of unlikely events and underestimate the occurrence of fairly certain events suggests a rational basis for what appears to be foolhardy driving behavior such as passing on a crowded hill or racing locomotives to grade crossings. Although the decision theory in these illustrations suggests plausible *a posteriori* explanations, it does seem reasonable to believe that specially formulated research to test its application to the intellectual performance of drivers would be appropriate.

The drivers' perceptual processes have usually been investigated in an effort to determine correct highway signing or similar visual requirements. It would seem worthwhile to depart from these relatively limited objectives to study the visual cues that are used by drivers in their task and to determine the extent that they are critical to him. These investigations may profitably be based upon the study of driver eye movements. In much the same way as earlier eye movement studies proved of value in instrument flight investigations it seems reasonable to assume that similar value could accrue to driver research. It is particularly fortunate that recent methodological and technical advances make the recording of eye movement superimposed upon the dynamic visual field a real possibility.[3] Here, at last, is a clear cut method for determining the extent of distraction of billboards. Or the effect of sensory saturation during long vigils at the wheel while driving over super highways. From this observer's viewpoint, the potential of such an approach is limited only by the imagination of the investigator.

In this discussion our concern has been with quantitative performance measures of system and driver. However, there is a clear requirement to consider driver preference as well as performance. Quite frequently human factors research has assumed that performance and preference lie on the same dimension. That is, people prefer the tool or machine with which

they perform best. On the basis of recent research it seems that this assumption must be modified to include the case where preference is measured along one dimension and performance on another. If the optima of these two dimensions indicate conflicting conclusions a problem of weighting the results arises. When an operator performs best during test with system A, but prefers working with system B there is no unequivocal method for determining to what extent either of the results should be compromised.

The manner in which preference judgments are collected may have considerable effect upon the results. We have found in an investigation of various control system dimensions that a straightforward application of the psychophysical method of constant stimuli has been effective as a model for collecting preference data and performance data under the same experimental conditions. Experimentally, the subject drove the test course using a conventional steering system and indicated his preference on a check list rating scale. Driving the same course with the experimental system, performance data were collected and the same check list was completed. This procedure was followed for all experimental conditions with precise and reliable results.

TABLE I

Matrix of Intercorrelations Among Dependent Variables

	$\int\|V\|$	$\int V^2$	$\int\|D\|$	$\int D^2$	$p = 0.50$	*Rating*
$\int\|V\|$	X	0.83	0.62	0.86	0.43	0.25
$\int V^2$		X	0.59	0.75	0.51	0.22
$\int\|D\|$			X	0.72	0.69	0.25
$\int D^2$				X	0.73	0.28
$p = 0.50$					X	0.28
Rating						X

The intercorrelations of the preference and performance measures are shown in Table I. The performance indices were measures of steering wheel movement. The velocity (V) and displacement (D) measures were related to driver response frequency, and the probability measure (p) is related to the probability of exceeding 0.5 radians/sec in steering-wheel movement amplitude. It is evident that the relationship between preference and performance is rather low. However, the reliability of the ratings was quite high ($r = 0.91$) indicating that the performance measures used were probably not sufficiently discriminating among the various experimental systems. This further illustrates the requirement for an overall system performance measure that would correlate with driver preference. Furthermore, it substantiates the possibility of simultaneously collecting preference and performance measures in experimental situations. In other experimental situations different judgmental methods might best be used. Regardless of what method is used there is a clear need for simultaneous collection of preference judgments and performance measures with man-operated systems and some quantitative method for combining them into a single criterion.

The general approach in this paper has been to illustrate the way in which a system measure of driver–vehicle performance might be utilized. The additional stress upon the measurement of human performance and preference is included because of the importance of isolating error sources. Error isolation is simply the process of identifying which component in a system accounts for the greatest proportion of variation in the output of the total system. It is of course apparent that the greatest overall increase in system performance can be achieved by improving the component that is the source of greatest error.

The measure of system performance that has been illustrated has implications for application beyond the automotive field. This approach may be useful to measure system performance wherever the following conditions prevail: (a) servo driven control surfaces are available that may be disturbed independent of the operator's control; (b) where the operator is instructed to perform a specific task, and (c) where the system motion equation may be isolated into yaw, roll, or pitch elements. These conditions

seem to be satisfied by most dynamic control vehicles from aircraft to submarines. With specific reference to a two- or three-dimensional system the operator might be instructed to maintain a constant heading, altitude or depth in the face of some disturbing force such as air gusts or sea currents. Hence the mean heading error of an aircraft or submarine is the triple integral of its yaw rate, and the mean altitude (or depth) error is the triple integral of its pitch rate. In these circumstances all of the advantages of flexibility, simplicity and validity should prove to be of as much value as they are in measuring automobile system performance.

REFERENCES

1. Fitts, P. M., Bennett, W. F. and Bahrick, H. P., Application of Auto-Correlation and Cross-correlation Analysis to the Study of Tracking Behavior. AFPTRC–TN–56–9, Jan., 1956.

2. McFarland, R. A., Moore, R. C. and Warren, A. B., *Human Variables in Motor Vehicle Accidents*. Harvard School of Public Health, 1955.

3. Mackworth, J. F. and Mackworth, N. H., Eye Fixations Recorded on Changing Visual Scenes by the Television Eye Marker. MRC/APU 320/58, Applied Psychology Research Unit, Cambridge, England.

A Re-evaluation of Machine-Generated Abstracts,

ALBERT RESNICK and TERRY R. SAVAGE, IBM Research Center and IBM Advanced Systems Development Division, Yorktown Heights, New York.

Twenty-five subjects were divided into five matched groups and each group was given a different type of lexical indicator of content. The purpose was to determine their ability to answer a 72-question criterion test and determine the relevance of the documents furnished relative to that task.

The results indicated that there were no major differences between the five groups in their ability to select relevant documents. The group utilizing complete text material, however, obtained a significantly higher score on the examination.

A previous experiment [1] compared four types of lexical indicators of content. The purpose was to determine (1) the feasibility of using machine generated abstracts to answer questions in a limited time period as opposed to using other types of lexical indicators, e.g. whole text or titles, and (2) which type of lexical indicators best permits humans to determine whether or not a document is relevant for a specific purpose.

At the conclusion of the prior experiment, it appeared that the results may have been strongly influenced by the type of subject population utilized and the artificiality of the task for such a subject population, and different results might have been obtained using a more homogeneous subject population and a criterion test with greater face validity.

These two features have been incorporated into the present experiment.

METHOD

The population of documents chosen for this experiment was formed by 75 reports selected from military intelligence documents which covered a wide variety of topics. Fifteen of the 75 documents were designated to form a subset of documents relevant to a criterion test. They were chosen so that they formed a more homogeneous group than the total document population. Two members of the intelligence staff, who were not familiar with the various types of machine abstracts, were asked to write 72 short-answer type questions utilizing the 15 relevant documents. They were asked to write questions that intelligence analysts could be expected to answer utilizing these documents. The questions were later edited by one of the authors.

The examination was constructed by randomly assigning 7 or 8 questions to each of 10 pages. The pages were then assembled in 25 different random orders.

Five sets of materials (T, R, A, P and C) taken from the total document population were prepared. Set T consisted of the titles only; Set R consisted of a random selection of sentences from each of the documents. Set A consisted of automatic abstracts prepared in the manner described by Savage, Fallon and Saxon[2]. Set P consisted of another type of abstract (pseudo-auto-abstract) formed by picking the same number of sentences from the documents as the auto-abstracting method (Set A), i.e. 10 per cent, but the particular sentences selected were simply the first 5 per cent and the last 5 per cent of the articles. Set C consisted of complete texts. None of the material contained captions or illustrations.

The titles and text were prepared using an IBM 407 while the three types of abstracts were prepared on an IBM 704. Thus, all of the material was printed in capital letters.

The design of the experiment is presented in Table I. The study was conducted utilizing 25 intelligence analysts as subjects. In the first of four experimental sessions, they were given the Robinson–Hall Geology Test to determine

1 Hugh Fallon and Marjorie Saxon did the programming for the generation of the material, ran subjects and helped tabulate the data. Carol Mattson and Al Barnes aided in the analysis of the data. This work was supported, in part, by the Department of the Army Contract No. DA-49-083 OSA-2109.

TABLE I

Experimental Design*

	Groups				
	T	*R*	*A*	*P*	*C*
Day 1	Robinson–Hall speed of reading test in geology and the sample criterion test	Robinson–Hall speed of reading test in geology and the sample criterion test	Robinson–Hall speed of reading test in geology and the sample criterion test	Robinson–Hall speed of reading test in geology and the sample criterion test	Robinson–Hall speed of reading test in geology and the sample criterion test
Day 2	Pre-test	Pre-test	Pre-test	Pre-test	Pre-test
Day 3	Titles	Random abstracts	Auto-abstracts	Pseudo-auto-abstracts	Text
Day 4	Titles plus text	Random abstracts plus text	Auto-abstracts plus text	Pseudo-auto-abstracts plus text	Text

* The 72-item questionnaire was administered to all five groups on Days 2, 3 and 4.

their speed of reading and comprehension. In addition, the subjects were given an examination composed of 15 questions similar in nature to the aforementioned 72 question criterion test but drawn from a different document population. They were to answer as many of these questions as they could in 55 min using eight documents, four of which contained answers to the 15 questions. The subjects were then divided into 5 groups which were matched on three variables: (1) the number of questions attempted on the sample criterion test, (2) the number of questions scored right on the criterion test, and (3) their speed of reading score on the Robinson–Hall Test. Table II shows the means and standard deviations of these measures for the total population and for each of the five groups.

In the second experimental session, all subjects were given the 72 question criterion test. They were instructed that this was a pretest to determine if they could answer any of the questions without any material to aid them. They were to answer as many questions as they could. They were told that in succeeding experimental sessions, they would be given material which would aid them in answering questions.

On the third day, three groups (R, A and P), which were given the random abstracts, auto-abstracts and the pseudo-auto-abstracts, reported together in the same room and the other two groups (T) and (C) reported to separate rooms. The abstract groups were told to imagine that their supervisor had asked them to answer a list of 72 questions. As they had a definite time limit, they would try to answer as many of the questions on the basis of what they already knew. In addition, they had a library which had given them a set of materials to aid them in this task. Group C was given the entire set of 75 texts on the third day and instructed similarly. Group T was given only the 75 titles on the third day. In addition to answering the questions, the subjects were asked to mark each of the documents with a "+" (acceptance), if they thought the document would aid them in answering the questionnaire, a "?", if they thought the document might aid them in answering the questionnaire, and a "–" (rejection), if they thought the document

TABLE II

Description of S Population on the Basis of which Groups were Matched

	Total population		*Group population*									
			T		*R*		*A*		*P*		*C*	
	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
No. tried on sample criterion test	12.8	1.4	13.0	1.3	12.4	1.6	13.0	1.3	13.0	1.3	12.8	1.5
No. right on sample criterion test	11.4	1.4	11.8	1.2	11.0	1.6	11.2	0.8	11.6	1.6	11.2	1.5
Reading rate R–H geology test†	241.9*	52.5*	258.4	34.0	225.4	65.0	244.8	38.6	249.2	48.6	231.8	61.9

* Words/min.
† Robinson–Hall.

would not aid them in answering the questionnaire.

On the fourth day, all groups were given the complete texts in addition to the materials they had on the third day. The purpose was to see if the abstracts and titles had been useful to the subjects in determining which of the complete texts should be consulted for answers.

Each of the experimental sessions were two and one-half hours long. Because of the quantity of material to be read, the "Text" (C) Group reading at the average rate of the 25 subjects could only complete one-quarter of the documents in the two and one-half hour experimental session on Day 3, assuming that they spent all of their time reading. The abstract groups could have completed at least two readings of each document in the allotted time period on Day 3.

At the end of the second and third sessions, the test papers were graded with a *C* for correct and an *X* for incorrect for each question answered.

On the third and fourth days, the subjects worked with the corrected questionnaire and were allowed to correct their previous incorrect answers but not erase them.

RESULTS

The subject's choice in the evaluation of useful documents can be categorized into correct acceptances, correct rejections, the acceptance of useless documents (Type II errors), and the rejection of useful documents (Type I errors). In addition, some of the documents were also marked with a "?" mark or were unmarked. Usefulness is here defined as a document which contained answer(s) to one or more of the 72 questions. Table III indicates, for the population in the five experimental groups, the percentage of the documents evaluated following each category of evaluation with the questioned and accepted categories combined and the rejected and unmarked categories combined. These categories were combined because it was obvious from the results that the unmarked documents had been summarily scanned and rejected and it is assumed that in any real system questionable documents will be looked at further.

TABLE III

Per cent Response of Subjects in Classifying Relevance of Document Population

	Groups on Day 2				
	T	*R*	*A*	*P*	*C*
Per cent of relevant documents*					
Accepted or questioned	79	79	83	79	76
Rejected or unmarked	21	21	17	21	24
	100	100	100	100	100
Per cent of irrelevant documents†					
Accepted or questioned	9	11	6	6	4
Rejected or unmarked	91	89	94	94	96
	100	100	100	100	100

* Relevant documents composed 20 per cent of the total document population.
† Irrelevant documents composed 80 per cent of the total document population.

A measure of the subjects' confidence in the usefulness of the material which they used on the third day is calculated from the ratio of acceptance (+), and rejection (−) responses to all the evaluative responses (+, −, and ?), i.e. the percentage of documents accepted and rejected over all the documents "scanned". This measure of confidence was high and varied little over groups. The lowest score was the "Title" (T) Group, at 89 per cent, and the highest was found for the "Auto-Abstract" (A) Group at 94 per cent. These differences, however, were not significant at the 0.05 level. Table III shows that for the combined categories there was relatively little difference among the five groups in determining the relevant and irrelevant documents.

Considering the questions answered by each group, on the pretest (Day 2), all five groups obtained about the same number of correct answers out of the total number of answers attempted which would seem to imply that the groups were evenly matched. On Day 3, when all five groups were given a different type of lexical indicator of content, the Text (C) Group showed the greatest improvement in correctly answering the 72 question criterion test over Day 2 as shown in Fig. 1. The three abstract groups exhibited intermediate improvement with the "Pseudo-Auto-Abstract" (P) Group doing significantly better than either the Auto-Abstract (A) Group or the "Random-Abstract" (R) Group. The difference between the scores of the Random-Abstract (R) Group and the Auto-Abstract (A) Group, however, were not statistically significant. On Day 4, when all five groups were given the complete text material, the Auto-Abstract (A) Group exhibited the greatest improvement in performance over Day 3, but the differences among groups in the number of correct responses made on Day 4, minus the number of correct responses on Day 2 plus Day 3, were not significant at the 0.05 level. These results are illustrated in Fig. 1 with an analysis of variance presented in Table IV.

In looking at the total number of responses right and wrong on Day 3 minus the total number of responses on Day 2, the differences between groups were not significant, except between the Text (C) Group and the other four groups and between the Title (T) Group and the Pseudo-Auto-Abstract (P) Group ($P = 0.01$).

DISCUSSION

The results point out that the subjects in the text group were able to determine the "usefulness" of documents and answer questions

based on them at rates faster than would be possible on the basis of straight reading of the texts. This shows that some selection process other than reading the entire document was employed by these subjects to determine relevant from irrelevant documents.

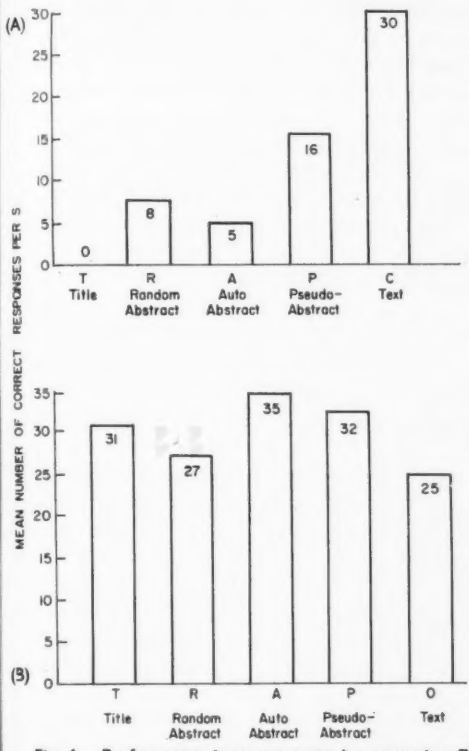

Fig. 1. Performance improvements in answering 72 short-answer questions. (A) Day 3 performance after subtracting correct pretest responses. (B) Day 4 performance after subtracting Day 3 performance after all five groups had been given text in addition to their previous material.

In the previous experiment, the use of titles to determine the relevance of documents to a criterion test without accompanying abstracts led to a high percentage of Type II errors, i.e. accepting documents which should be rejected. This was not true in this experiment which may have been due to the fact that the titles of the relevant documents in this experiment were more indicative of the contents of these documents than in the previous experiment. The document population in the earlier study, unlike the document population in this study, had less variety in subject matter involving primarily the biological and medical areas with many of the irrelevant titles appearing relevant.

TABLE IV

Analysis of Variance

Total responses:	*df*	*Mean square*	*F*
Pretest			
Between groups	4	37.1	0.8
Within groups	20	43.9	
Day 3			
Between groups	4	655.2	0.9
Within groups	20	693.9	
Day 3 minus Day 2			
Between groups	4	625.2	16.5*
Within groups	20	37.9	
Day 4			
Between groups	4	368.9	5.6*
Within groups	20	65.7	
Day 4 minus Day 3			
Between groups	4	78.9	1.6
Within groups	20	49.1	
Correct responses:			
Pretest			
Between groups	4	0.8	0.2
Within groups	20	3.8	
Day 3			
Between groups	4	697.5	56.2*
Within groups	20	12.4	
Day 3 minus Day 2			
Between groups	4	696.5	54.4*
Within groups	20	12.8	
Day 4			
Between groups	4	486.7	7.3*
Within groups	20	66.8	
Day 4 minus Day 3			
Between groups	4	74.1	1.5
Within groups	20	48.9	

* Significant at $P < 0.01$.

In looking at the number of answers which were contained in the various types of lexical indicators, it was found that the auto-abstracts contained 26 per cent of the answers; the pseudo-auto-abstracts, 25 per cent; while the random abstracts contained 15 per cent and the titles 0 per cent. The text, of course, contained all the answers. The Text (C) Group was apparently hindered by the volume of material to be read or scanned as they were only able to find 45 per cent of the available answers on Day 3. While the Random-Abstract (R) Group found 69 per cent of their available answers and the Pseudo-Auto-Abstract (P) Group 86 per cent, the Auto-Abstract (A) Group found only 25 per cent. The Auto-Abstract (A) Group should have been able to answer more questions correctly than either of the other two abstract groups considering the number of answers contained in their material, but they did not. No explanation could be found for this result.

One would expect that the use of abstracts and titles on Day 3 might have aided these groups in locating answers on the fourth day when they had access to the complete texts. Indeed, the abstract and the title groups did slightly better than the text group on the fourth day. This difference, however, was not statistically significant and is confounded with the fact that the text group had fewer questions left to answer on the fourth day. It seems, therefore, that exposure to the abstracts was not of much value in using the complete texts.

On the basis of the results, it would appear that, as in the previous experiment, texts and abstracts are better than titles for the purpose of obtaining answers to specific questions; the complete text being better than abstracts for this purpose even with the time limitations imposed by this experiment. But, if the task is to determine whether documents are relevant or not, all five types of lexical indicators seem to be about equal to each other under the experimental conditions in this study.

SUMMARY

Twenty-five subjects were divided into five groups matched for their reading speed and scores on a sample criterion test. Each group was given a different kind of lexical indicator of content for a set of 75 documents, namely, titles, three types of abstracts and complete texts. In addition, they were given an examination composed of 72 short-answer type questions which were derived from 15 out of the 75 documents. They attempted to answer the questions and evaluate the relevance of each document in answering the examination.

There were no major differences between the five experimental groups in their ability to select documents relevant to the criterion examination. The Text (C) Group, however, obtained a significantly higher score on the examination.

In comparing these results with an earlier study, it was found that similar results were observed in that the text group out-performed the other groups in obtaining correct answers; but in the previous study, the text and abstract groups out-performed the title group in determining the relevance of documents to the criterion test, whereas in the present study all groups did equally well.

REFERENCES

1. Rath, G. J., Resnick, A. and Savage, T. R., Comparisons of Four Types of Lexical Indicators of Content, RC 187, Aug., 1959. (Copies may be obtained by writing any of the authors at IBM Research Center, Box 218, Yorktown Heights, New York.)

2. Savage, T. R., Fallon, H. C. and Saxon, M. E., Acsi-Matic Auto-Abstracting Project, Interim Report, p. 27. IBM Advanced Systems Development Division; Aug., 1959.

Dynamic Anthropometry of Working Positions

H. T. E. HERTZBERG, Aerospace Medical Laboratory, Wright Air Development Division.

This paper provides a review of the principles and procedures of workspace design for engineers. It emphasizes that human body size, anthropometry, and muscle force capability, biomechanics, are both essential for the efficient sizing of equipment. The proper method of workspace design, the "design limits concept", is described, and the fallacy of the "average man" concept is demonstrated. General methods of gathering body size and strength data are outlined, and major information sources noted. The author's ideas on human muscle strength in the weightless state are included.

In workspace design, it is axiomatic that both human body size and muscle "strength" are prime factors: first the man must be able to reach the controls, then be able to operate them in order to make the machine do his work. Designers have developed efficient engineering formulas to minimize the amounts of materials for best construction, and teach such formulas thoroughly to students in all engineering schools. But the formulas to achieve the highest integration between the machine and its operator—which, after all, is the final payoff of design—are not strictly engineering formulas, derived mathematically by engineers; they are rather human engineering formulas or principles developed by anatomists, anthropologists, physiologists, psychologists and other human factors people. Reliable anthropometric data and procedures form the most powerful tool available today for the optimal sizing of many mass-produced items, whether truck-cabs or oxygen masks. Very few engineering schools teach these principles in their entirety, and most of them not at all. It is a matter of experience that many young graduate engineers are unacquainted with the use of human factors data, and that on innumerable occasions designers still size their work according to the fallacious concept of the "average man". Furthermore, many mature engineers, usually not in the aviation industry, appear largely unaware of the great advances achieved since World War II in the accommodation of human beings in aircraft and other confined workplaces. Unfortunately some of these people are in positions of power to make far-reaching decisions. In this discussion, therefore, I shall try to present a bird's-eye[1] introduction to the principles, data and sources comprising the anthropological aspects of human engineering.

But before discussing any subject, it is always desirable to define it and the basic terms to be used. Our title mentions "Anthropometry", but it is well to place this word briefly in its proper context. "Anthropology" is the overall study of man as an animal and as a social being; "Physical Anthropology" is the sub-discipline concentrating on man as an animal, including the measurement of differences in dimension among living varieties, and "Anthropometry", meaning literally the measurement of man, denotes the techniques of body measurement and the tabulated data themselves. It also implies the statistical approach, the measuring of an adequate sample of men to permit an orderly arrangement of data from the smallest to the largest. For standardization, however, conventional anthropometric dimensions have to be taken on subjects in static poses; whereas anyone who does work needs room to move in. The broader term, "Dynamic Anthropometry", thus usually implies the study of the workspace of space envelope needed by men as they perform their work. Yet this term can also carry the implication of an evolving science of workspace, and I shall use the term in both senses hereafter.

Sound workspace design involves both

[1] It has to be bird's-eye, because at present there are no courses or textbooks to synthesize the 1000-odd titles setting forth the facts that are important to Applied Physical Anthropology.

objective factors (body size, muscle force capability) and subjective factors (individual preferences in types of seating surfaces, pain, fatigue, etc.), all of which must be understood to insure long-term occupant comfort, but I shall outline here only some of the objective factors of primary importance to engineers. They are the first steps essential to fully adequate workspace design. The subjective factors are perhaps better left to other human factors people.[2]

The dimensions needed for workspace vary with the man's duties, but most duties are performed sitting down. In planning a workspace for a seated man, the designer must know first the occupant's static dimensions (sitting height, buttock–knee length, knee height, arm reach, etc.) and then his dynamic dimensions, which include both the distances the man can move and forces he can exert. If only one occupant, is to use the workplace, the designer's task is easy—he tailors it to the subject. But suppose many persons must work efficiently within it? Shall the designer then size it for the "average man"?[3]

This is a solution frequently adopted, but as a basis for design it is a fallacy. There are persons who are average in sitting height, or in buttock–knee length, but no one is average in all dimensions. Let us look at the findings of a study made some years ago by Daniels and Churchill[2], to test the assumptions of the "average man" concept. They examined the body size data on more than 4000 flying personnel to see how many men could be average in 10 dimensions useful in clothing design. But instead of using merely the exact average, they generously included plus and minus 15 per cent from the average in their figure, or the middle 30 per cent of the sample for each dimension. They called these "approximate averages", and used the approximate average of each of the dimensions as hurdles in a step by step elimination.

The following passage is taken from their report. Only the percentage table has been added.

	Per cent of original sample
(1) Of the original 4063 men, 1055 were of approximately average stature.	25.9
(2) Of these 1055 men, 302 were also of approximately average chest circumference.	7.4
(3) Of these 302 men, 143 were also of approximately average sleeve length.	3.5
(4) Of these 143 men, 73 were also of approximately average crotch height.	1.8
(5) Of these 73 men, 28 were also of approximately average torso circumference.	0.69
(6) Of these 28 men, 12 were also of approximately average hip circumference.	0.29
(7) Of these 12 men, 6 were also of approximately average waist circumference.	0.14
(8) Of these 6 men, 3 were also of approximately average waist circumference.	0.07
(9) Of these 3 men, 2 were also of approximately average thigh circumference.	0.04
(10) Of these 2 men, 0 were also of approximately average crotch length.	0.00

This list shows that even the "approximately average" man in only 1 dimension is already a relatively uncommon man (about 1 in 4), while less than 8 in 100 show the average in only two dimensions; and for four dimensions, the number for all practical design purposes has shrunk to the vanishing point in the population.

The fact that these figures include an exceedingly generous 30 per cent of the population in

---

[2] Engineers do not usually concern themselves with subjective factors, but it should be noted that two teachers in an engineering school produced a classic study of human comfort.[8]

[3] For the sections on the "average man" and "design limits", I have drawn heavily on a previous paper.[11]

each dimension makes the demonstration even more convincing. It should be emphasized that any other set of dimensions, including those for workspace, would show the same results.

Thus we cannot escape the conclusion that the "average man" concept is not a sound basis for design. The only efficient method is the concept of "design limits", the anthropometric–statistical approach based on sound dimensions measured by standardized techniques on the population that will use the workspace.

The "design limits" concept is not the product of any one person. It is the result of many men's experience, both in the United States and abroad. So far as I know, Randall, Damon and the other anthropologists at Wright Field during World War II were the first in the United States to apply the principles in a broad way, beginning in 1942, and to summarize the work of years in a major publication.[22] Since those beginnings the realization has grown with commentary by Morant [17]; Hooton [13]; King, Morrow and Vollmer [14]; Morant [18]; Morant and Ruffell-Smith [19]; Weddell and Darcus [26]; Roebuck [23]; McFarland, Damon and others [16]; Hertzberg and Saul [10]; Hertzberg [11]; and Ely and his associates [5,6], to name but a few. Among human factors workers this concept is universally accepted. But because the "average man" concept continues to be used by many designers, a review of the "design limits" concept appears necessary.

In its essentials, this system, sometimes also called the "range of accommodation" system, is very simple. Take relevant dimensions on an adequate sample of the using population, and reduce the date statistically; then from these data choose the dimensions for workspace according to the percentage of the population that the product is intended to accommodate.

One of the simplest examples of the "design limits" concept is the fighter cockpit, whose plastic canopy protrudes just high enough to cover the pilot's head. Obviously the pilot's eye level, regardless of his body size, must be approximately constant and therefore the pilot's seat must be vertically adjustable. The anthropometric dimension, Sitting Height, consequently becomes one of the major determinants of cockpit size.

If we measure a large random Air Force sample for Sitting Height, we can list each man's dimension in numerical order to find the range of variation. We get percentiles by dividing the range into 100 equal steps. For example, the 95th percentile of any dimension is that value which is large enough to include 95 per cent of the population, but still is below the remaining largest 5 per cent.

TABLE I

Percentile Values for Sitting Height of Air Force Flying Personnel (1950)

Percentile	*In.*	*Percentile*	*In.*
0	29.9	60	36.3
1	32.9	70	36.6
2	33.3	80	37.0
3	33.5	90	37.6
5	33.8	95	38.0
10	34.3	97	38.3
20	34.9	98	38.6
30	35.9	99	38.9
40	35.6	100	41.2
50	36.0		

Table I shows the range of values of Sitting Height found to exist in a sample of 4061 Air Force flying personnel in 1950.[9]

If we now plot percentile versus dimension, we obtain the curve in Fig. 1. We see immediately that there is a straight-line portion of moderate slope comprising approximately the central 90 per cent of the sample, the remaining 10 per cent divided into two zones of sharply increased slope.

Now the extremes of any curve like this are unstable, depending on the appearance in sample of a few very small or very large individuals. To reduce such instability, statisticians usually use the mean ± 3 standard deviations,[4] or 99.7

[4] A standard deviation is a number indicating the probable inherent variability of a sample about its mean value.

per cent of the range. One of the beauties of the percentile curve is that the design engineer does not need to be a statistician to use it; he need only choose the design limits he wants. He may find it simpler to deal with the 98 per cent falling between the 1st and 99th percentiles (in this case, about 2.3 standard deviations).

Perhaps the outstanding fact here is that the central 90 per cent of the sample shows only about 40 per cent of the total variation, while the extreme 10 per cent of the sample shows about 60 per cent of the total variation. We note also that the combined variabilities of the extreme 2 per cent practically equal that of the

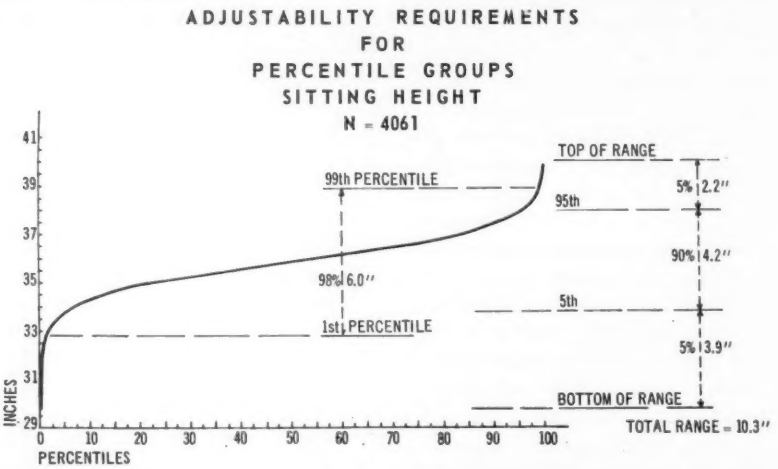

Fig. 1. The percentile curve for sitting height in a USAF population, showing seat adjustability requirements for various percentile groups (adapted from Hertzberg[11]).

In this curve, the variability of the total sample is seen to be 10.3 in. The variability of the central 98 per cent drops to only 6.0 in., and the variability of the central 90 per cent is only 4.2 in. Table II summarizes these variabilities, and presents the adjustabilities required for each percentile group in a fighter cockpit, should we have to accommodate the full sample.

TABLE II

Per cent of sample ($N = 4061$)	*Variability* (in.)	*Adjustability required* (per cent)	
100 per cent	10.3	100	
Central 90 per cent	4.2	40.8	
Largest 5 per cent	2.2	21.3	59.2
Smallest 5 per cent	3.9	37.9	
Largest 1 per cent	1.4	13.6	42.7
Smallest 1 per cent	3.0	29.1	

central 90 per cent. From such data the engineer can judge the relative expense of accommodating different segments of the population. The data explain the general rule that, in the design of U.S. Air Force cockpits, flying clothing or personal equipment, each must accommodate at least 90 per cent of flying personnel, but that only certain types of equipment need to do more. Fighter cockpits may be barred to men above the 95th percentile in size; but oxygen masks should fit 100 per cent of the flying population.

Curves for other body dimensions can be used in similar fashion. The designer must know when to use the maximum or minimum value of a dimension. For example, in designing a table and chair combination, a maximum value of Knee Height, Sitting, should be chosen to determine the height of the bottom edge of the knee well, so as to accommodate the largest

man who is to use it. The height of the adjustable chair, however, should start from the minimal value of Popliteal Height[5] and extend through the chosen range of that value. If women are to use the chair as well as men, the minimal height should be the lowest value for women. By the same token, if the chair occupant must also see over instrument panels on the desk, the minimal value of Eye Height, Sitting[5] should limit the height of such panels or obstructions. These and other principles are detailed in[1].[6] If other working positions are needed, such as standing, crawling or kneeling, data for some of these are available.[12]

It should be remembered that anthropometric data are taken on subjects who are nude or nearly so. The addition of ordinary light clothing makes no important dimensional difference; but if the man must wear Arctic clothing, a suitable increase in workspace must be provided. This of course includes more space for booted feet, greater spacing between switches for gloved fingers, larger seats and a whole gamut of other increases over nude dimensions. Therefore the designer must keep in mind the environment as well as the man. Table III presents increments needed for many ordinary purposes.

We can now begin to appreciate the great power inherent in reliable anthropometric data. With facts like these the engineer can size his workspace to the segment of the population he chooses, or, conversely, if he is unavoidably limited for some valid reason to a given size, he can ascertain the population segment his product will accommodate.

The anthropometric dimensions discussed above have generally been taken with standard instruments—anthropometer, calipers and tape—but these are not the only way of getting workspace information. Swearingen[25] has devised a clever workspace measuring device in which the man sits at the center of what resembles a large pin cushion. He is surrounded by slender shafts or staves which can be moved radially toward him until he can touch the tips of each. These points in space describe his work envelope. But this device, though very effective, requires a whole room and considerable structure to house it. Dempsey[3] has somewhat simplified Swearingen's arrangement by reducing the number of staves necessary. Still other variations of this theme are possible.

TABLE III

Approximate Amounts Added to Nude Dimensions by Heavy Winter Flying Clothing*

	In.
Stature	1.9
Sitting Height	0.6
Buttock–Knee Length	0.5
Knee Height, Sitting	1.8
Shoulder Breadth (Bideltoid)	0.7
Waist Depth	1.4
Hip Breadth, Sitting	1.7
Hand Breadth	0.4
Foot Length	2.7
Foot Breadth	1.2
Elbow Breadth	4.4

* These increments are for certain types of Arctic Flying Clothing.[22] While civilian clothing will not necessarily be the same—the foot size increase particularly reflecting the use of special boots not normally available to civilian populations—the increases are generally reasonable and in the right direction.

Photography, of course, has been used by many workers as an effective means of recording workspace data. Probably the most complete of such systems is that of Morant[18] who photographs a man sitting in the corner of a room with his back touching one wall and his side the other. The photographs are from both front and side, and grid scales on the wall provide the dimensions. Still other photographic methods[9] are possible.

But all of these methods—anthropometric, mechanical, and photographic—provide essentially static dimensions, whereas dynamic dimensions are also needed for adequate workspaces. This is not to decry static dimensions;

---

5 These and other names of dimensions are taken from reference 9.

6 Designers for most U.S. applications would be well advised to rely more on recent USAF or OQMG data in applying these principles than upon the British or medical sources cited.

they constitute the bulk of present anthropometric data. Excellent results have been achieved by their use, and engineers can effect many further workspace improvements without exhausting present data (the sources of which will be outlined later). But the man must move and exert force in his space envelope, and this brings us to Biomechanics.

Biomechanics is the term frequently used to describe the physical properties of the human body as a dynamic solid—mobility, muscle force capability, centers of gravity of segments and of the body as a whole, masses of segments and many others. Here our concern will be only muscle force capability and mobility.

Numerous workspace guides exist (various Air Force Manuals),[7] and, more recently, reference[6] that specify the maximum forces to which levers and controls are to be designed, which is of course sound practice. But muscle force output of a limb is not constant; it varies with limb position, and there is a position of each limb at which its force output is at a peak. An example from an unpublished study may serve as an example.

A cockpit mock-up was constructed to study the torques that could be applied by the foot upon a brake pedal. The brake pedal could be set at various successive angles in the arc of foot rotation. The rudder bar itself could be modified from neutral leg position to extended leg position. Muscle forces of the foot rotating about the ankle—not full leg forces on the rudder bar—were measured (by means of strain gauges on the brake pedal) for both neutral and extended leg at 11 positions of the pedal (5° to 55° past vertical) throughout the angular range of foot movement. All of these measurements were taken on each subject in 3 different cockpit sizes,[8] 37-in., 39-in., and 41-in. In Fig. 2, the summarized pressures of 100 pilots have been plotted on a polar graph. The results demonstrate that the greatest muscle force for all 3 cockpit sizes and for both leg positions is produced by the foot when the pedal is between about 20° and 35° forward of vertical. This, then, is the foot position for most efficient braking action, when braking is a foot operation, rather than a leg thrust.

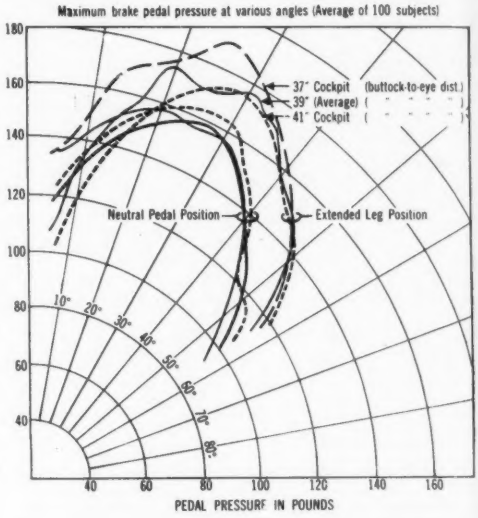

Fig. 2. Maximum brake pedal pressure at various angles (average of 100 subjects).

Thus the study of muscle force is not merely a matter of ascertaining maximum forces men can exert; it is rather a matter of locating within the workspace those zones where men can naturally exert the greatest force; in other words, where force is exerted most efficiently by the body. Controls designed to take advantage of the natural capability of the body should be placed in those positions.

7 HIAD—Handbook of Instructions to Aircraft Designers (ARDCM 81-1)
HIGED—Handbook of Instructions to Ground Equipment Designers (ARDCM 80-5)
HIAGSED—Handbook of Instructions for Aircraft Ground Support Equipment Designers (ARDCM 80-6)
HEGED—*Human Engineering Guide to Equipment Design* (in preparation).

8 The size of USAF aircraft cockpits is designated according to the vertical dimension from the floor of the cockpit to horizontal line of vision of the occupant. The height of the seat reference point above the floor increases with increasing cockpit size.

Workspace combining both static and dynamic data can be quite effective, yet they do not fully provide for the complexities of the human work needs. The final refinement will be possible when both space requirements and force output are measured simultaneously on the man. Static measures here would be wholly inadequate; the proper tool is strobophotography. Dempster has used this system with mirrors to obtain a 3-dimensional effect. Drillis, in a series of papers [4] has developed the photographic technique to a very high degree of precision for the analysis of bodily movement. Thus here is a choice among numerous techniques for specific kinds of problems.

It is now largely a matter of getting a complex and time consuming set of tests run on statistically adequate numbers of representative subjects. It can be distinctly misleading, as has been emphasized, to use just one "average" man or even only a few men; not less than 20 men, carefully chosen to cover the known range of body size, or preferably about 50, should be used. Otherwise the precision of biomechanical measurements may be completely lost in the gross error of the sample.

These techniques and principles became even more useful in the design of workspace for space travelers. The cost of hoisting any manned capsule into space is so great that the enclosure must be as small as possible while yet providing adequate space for the range of human body sizes it must accommodate. Such efficiency can be achieved only by the methods outlined above.

A note on the biomechanics of weightlessness seems relevant here. The widely publicised conceptions of space engineers like von Braun and Stuhlinger [24] are so large that they would probably be prohibitively massive to hoist intact, on even keel, after fabrication on earth; hence the idea is to build them in space from sub-assemblies sent aloft piecemeal. Such construction, of course, means hand labor.

A man protected by a capsule can probably continue to exist in space, but he can do no construction work. Consequently the fabrication of stations in space depends first of all on the development of a suit in which the man can live and move, that is, work, for some hours at a time before he has to get back into the capsule that brought him there, either for return to earth or rest and replenishment of his supplies. Until such a suit is developed, construction of anything in space seems impossible.

Yet even when he has a suit, certain aspects of the space environment will govern his ability to work. First, there is no gravity, hence there is no traction. Therefore all methods of exerting muscle force depending on traction must be ruled out. This means, for example, that in outer space, riveting as we know it, where one man with a rivet gun exerts forces which are opposed by another man, would be impossible. Without traction, neither man could oppose the other, and the rivet gun forces would drive them away from their work. From this mundane example, it is clear that the biomechanics of weightlessness must be carefully considered in planning work in outer space. Second, the human body on earth is a prime mover that can exert forces in two ways. In the first, one limb may exert force directly on other limbs, which can be called a "closed" force system; or the limb may direct its force upon other objects external to the body, which might be called an "open" system.

In a closed system, the fingers might grip a ball, or the hands push against each other, or pull on opposite ends of a shaft. In an open system, the hand may hammer a nail, or a man may buck a rivet by leaning against it. In all open systems, either directly or remotely, the earth is a part of the force diagram, and traction due to gravity is the usual means whereby the earth can be used. Man's body with its whole method of locomotion has evolved according to this fact, and is conditioned to gravity and traction from birth. When we now place this organism in an environment that lacks gravity and traction, we are severely hampering its effectiveness, because so large a part of its conventional force output depends on gravity. (We may here neglect the fact that an appreciable

part of the alimentary processes—e.g. the act of drinking—on which man's energy depends, also make use of gravity.) Hence it is clear that some substitute for traction must be developed so as to create closed force systems in space to substitute for the numerous open systems that we habitually use on earth.

It seems rather likely that the muscle force that a pressure-suited man can apply in closed systems in space will not be severely reduced from what he can exert on earth. Of course, there will be an appreciable decrement due to the sheer restrictiveness of a pressure suit, the internal friction of the garment layers, the tendency of the inflated suit to assume and maintain a globular form, and the slippage and clumsiness that inevitably accompany the use of heavy gloves, but the decrement should not be so large that the man's closed-system strength is reduced to only a small fraction of its normal value. There is reason to hope that the decrement may not exceed about 50 per cent. In an open system, however, the force capability would be reduced nearly to zero. As Gauer and Haber [7] have pointed out, a weightless man exerting torque on a shaft will tend to rotate about the axis of shaft. Thus open force systems must be avoided, and all manual work to be performed in space must be planned in minute detail for the conditions to be encountered, or it will be impossible of accomplishment.

Fortunately, many expedients are available. If a space satellite is to be built in space, for example, it will have to be sent up as parts, ready for assembly. If ordinary riveting methods cannot serve, then a system of explosive rivets might be used. A man could carry a high-voltage battery on his back—it would not burden him with its weight—and, clipping the ground wire to the frame, touch the rivet with a probe from the hot lead, thereby exploding it. This method could permit the joining of many parts in a reasonably short time. Again, if a man must exert force on a surface, he could hold himself to it by means of a harness on his body with straps to the surface. If the surface were narrow enough, he could grip the edges with rollers at the ends of the straps, thereby enabling him to crawl or walk the length of the surface; otherwise lugs or slots or other means of attachment might well be used with quick release mechanisms for the same purpose. Magnetized shoes are possible—or even magnetized gloves to pull one's self along with—but iron is so heavy that its use may be ruled out except in limited areas.

In addition, the number of tools required must be at a minimum, and these must be tied to him, for he will have nowhere to lay them down.

Since workspace engineers or other designers can hardly be blamed for not using anthropometric data they do not know about, the U.S. Air Force has recently published its "Annotated Bibliography of Applied Physical Anthropology in Human Engineering".[8] This is a compendium of working data extracted from 121 reports and articles categorized by subject matter, title and author. It is unnecessary to recapitulate that bibliography here, but it may be emphasized that even this group of reports constitutes less than 10 per cent of the available literature. Suffice it to say that there is a wealth of anthropometric information, much of it generated by the Office of the Quartermaster, U.S. Army, and Wright Air Development Center, U.S. Air Force, which should be available to designers. Because many academic groups are also active in this field, professional journals of long standing like *American Journal of Physical Anthropology* and *Human Biology* in the United States, *Ergonomics* in Great Britain and *Arbeitsphysiologie* in Germany, should also be included in all engineering libraries. These last two in particular publicize much applied science that will interest the workspace designer. *Human Factors*, a journal now supplying this need in the United States, should be there also. Chapter IX of a joint publication of the Army, Navy and Air Force, "Human Engineering Guide to Equipment Design", is being prepared as a summary of the entire field of anthropological design principles and data, and should be available within about six months.

It is hoped that the engineering profession will see that such material is included in engineering curricula for the benefit of management and worker alike.

Perhaps the best summary of these remarks would be a brief paraphrase of remarks by Randall, Damon, and their collaborators [22]:

1. Man is functional,
2. Functional men vary in size,
3. Therefore consider the man, for he is a vital factor in design, and
4. Consider him early, so that his accommodations will be adequate, and will neither hamper his operation nor limit his size.

REFERENCES

1. Coakley, J. D., Fucigna, J. T. and Barmack, J. E., A Functional Application of Anthropometric Data to the Design of the Workspace of PPI Scope Operators. WADC TR 53-3, Wright Air Development Center, Wright–Patterson Air Force Base, Ohio, 1953.

2. Daniels, G. S. and Churchill, E., The "Average Man"? T.N. WCRD 53-7, Wright Air Development Center, Wright–Patterson Air Force Base, Ohio, 1952.

3. Dempsey, C. A., Development of a Workspace Measuring Device. WADC TR 53-53, Wright Air Development Center, Wright–Patterson Air Force Base, Ohio, 1953.

4. Drillis, R. J., The use of the gliding cyclograms in the bio-mechanical analysis of movements. *Human Factors*, **1** (2), 1, 1959.

5. Ely, J. H., Thompson, R. M. and Orlansky, J., Layout of Workplaces. Chap. V of the Joint Services Human Engineering Guide to Equipment Design. WADC TR 56-171, 1959.

6. Ely, J. H., Thompson, R. M. and Orlansky, J., Design of Controls. Chap. VI of the Joint Services Human Engineering Guide to Equipment Design. WADC TR 56-172. Wright Air Development Center, Wright–Patterson Air Force Base, Ohio, 1956.

7. Gauer, O., and Haber, H., *Man in Gravity-Free Conditions. German Aviation Medicine in World War II.* Vol. 1, 1950.

8. Hansen, R. and Cornog, D. Y., Annotated Bibliography of Applied Physical Anthropology in Human Engineering. Edited by H. T. E. Hertzberg. WADC TR 56-30, Wright Air Development Center, Wright–Patterson Air Force Base, Ohio, 1958.

9. Hertzberg, H. T. E., Daniels, G. S. and Churchill, E., Anthropometry of Flying Personnel—1950. WADC TR 52-321, Wright Air Development Center, Wright–Patterson Air Force Base, Ohio, 1954.

10. Hertzberg, H. T. E. and Saul, F. P., Anthropometry in Workspace Design. Section C of Chap. X (Human Factors in the Design of Bombing Systems), from *Introduction to the Theory and Practice of Bombing.* (Title and Section C are unclassified, report is secret.) Edited by Clemens and Johnstone. Raytheon Manufacturing Co., Waltham, Mass., 1955.

11. Hertzberg, H. T. E., Some contributions of applied physical anthropology to human engineering. *Ann. N.Y. Acad. Sci.* **64**, Art. 4, 1955.

12. Hertzberg, H. T. E., Emanuel, I. and Alexander, M., The Anthropometry of Working Positions: 1. A Preliminary Study. WADC TR 54-520, Wright Air Development Center, Wright–Patterson Air Force Base, Ohio 1956.

13. Hooton, E. A., *A survey in Seating.* Heywood-Wakefield, Gardner, Mass., 1945.

14. King, B. G., Morrow, D. J. and Vollmer, E. P., Cockpit Studies—the Boundaries of the Maximum Area for the Operation of Manual Controls. Project X-651, Report No. 3, Naval Med. Res. Inst., Natl. Naval Med. Center, 15 July, 1947.

15. Lay, W. E. and Fisher, L. C., Riding comfort and cushions. *S.A.E. Journal (Transactions)*, **47**, 482, 1940.

16. McFarland, R. A., Damon, A. *et al.*, *Human Body Size and Capabilities in the Design and Operation of Vehicular Equipment.* Harvard School of Public Health, Boston, Mass., 1953.

17. Morant, G. M., The Choice of Subjects Acting at Mock-Up Trials of Pilot's Cockpits and Gun-Turrets. Flying Personnel Research Committee, No. 596, RAF Physiology Lab., Air Ministry, Great Britain, 1940.

18. Morant, G. M., A Proposed Anthropometric Technique Suitable for Standardization and Intended to Aid the Solution of Problems Concerning the Dimensions of Cockpits and Other Crew Positions in Aircraft. Report of the Working Party on Bio-Anthropometrical Problems of the Aeromedical Panel of the Advisory Group for Aeronautical Research and Development (NATO), 1954.

19. Morant, G. M. and Ruffell-Smith, W/Cdr. H. P., Body Measurements of Pilots and Cockpit Dimensions. FPRC 689, RAF Inst. of Av. Med., Farnborough, Hants., Great Britain, 1947.

20. Patt, D. I. and Randall, F. C., Principles of Seating in Fighter Type Aircraft. Memorandum Report No. TSEAL 3-695-58, Air Matériel Command, Wright–Patterson Air Force Base, Dayton, O., 1945.

21. Patt, D. I. and Randall, F. C., Principles of Cockpit Seating. Memorandum Report No. TSEAA-695-58c, Air Matériel Command, Wright–Patterson Air Force Base, Dayton, O., 1946.

22. Randall, F. C., Damon, A., Benton, R. S. and Patt, D. I., Human Body Size in Military Aircraft and Personal Equipment. Army Air Forces TR 5501. Air Matériel Command, Wright–Patterson Air Force Base, Dayton, O., 1946.

23. Roebuck, J. A., Anthropometric Data: Human Body Dimensions and Weights. Report No. SM 14617, Douglas Aircraft Co., Santa Monica, Cal., 1952.

24. Stuhlinger, E., Outlook to space travel. *Sci. Mon., N.Y.* (December), **87** (1957).

25. Swearingen, J. J., Design and Construction of an Adjustable Cockpit Mock-Up. Project No. Biotechnology. Civil Aeronautics Medical Research Laboratory, 2-48, Report No. 1 (1949).

26. Weddell, A. G. M., and Darcus, H. D. Some anatomical problems in naval warfare. *Brit. J. Industr. Med.* **4**, 77, 1947.

Volumetric Requirements for Hand Tool Usage[1]

PAUL T. BAKER, JAMES M. McKENDRY and GEORGE GRANT, HRB-Singer, Inc., State College, Pennsylvania.

This paper presents a quantitative description of the volume required to use common hand tools. The volume described, termed the "space envelope", is for the tool using hand of right-handed subjects. The data were collected on six individuals who had hands at or above the 95th percentile of the military population in length and breadth. The data were obtained by use of a photographic process employing time exposures for each action. This technique provided data comparable to that provided by movie techniques but at a lower cost. The implications of the data for maintenance requirements are also discussed.

INTRODUCTION

Component accessibility has become a problem in the design of complex electronic equipment. Despite improved component technology, it is still necessary to design equipment so that the components can be reached and serviced easily by maintenance men. These requirements mean space; space which cannot be filled in such a manner as to prevent, or slow down, proper maintenance.

The published data fail to provide adequate guidance on the space required for maintenance tasks. Existing reports, such as that of Jones *et al.*[3], delimit the measurements of the static hand. However, liberal extrapolation is required to translate these measurements into estimates of the size of a hand complete with the desired tool. Furthermore, a maintenance action requires dynamic hand-tool movements which increase the space required by the hand and tool. In this study the space requirement has been termed the "space envelope" and attempts have been made to define the overall dimensions of this space envelope for basic one-hand maintenance actions.

1 This research was supported by the U.S. Navy under Contract N 61-339-330, monitored by the Aviation Psychology Division, Human Engineering Department, of the U.S. Naval Training Device Center, Port Washington, N.Y. The opinions and conclusions in this report are those of the authors. They are not to be construed as reflecting the views or endorsement of the Department of the Navy. Acknowledgement is made to other members of the HRB-Singer staff who assisted in the research. Dr. John F. Corso provided help during the early planning stages, and Mrs. Thelma Baker conducted the statistical evaluation.

PROCEDURE

Six individuals were studied performing the following specific actions:

(1 & 2) Removing and replacing a miniature tube (9-pin with $\frac{7}{8}$ in. diameter and $1\frac{15}{16}$ in. height).

(3 & 4) Removing and replacing an average tube (octal base $1\frac{9}{32}$ in. diameter and $2\frac{3}{4}$ in. height).

(5 & 6) Removing and replacing a large tube (octal base with $2\frac{1}{16}$ in. diameter and $4\frac{3}{4}$ in. height).

(7) Using a small screwdriver ($2\frac{3}{4}$ in. length) to turn a screw.

(8) Using a large screwdriver (8 in. length) to turn a screw.

(9) Using an average spintite wrench (6 in. length) to turn a nut.

(10) Using an average pair of pliers (needle-nosed with $5\frac{1}{2}$ in. length) to unwind a resistor wire.

(11) Using an average pair of wirecutters (5 in. length) to cut a wire.

(12) Using an average size Allen wrench (2 in. length) to remove a knob.

(13) Using an average-sized socket wrench ($\frac{3}{8}$ in. base with $3\frac{1}{4}$ in. shaft) to turn a nut.

Size was varied in cases where it was thought to affect the motion employed. When the motion seemed to be unaffected by size, it was not varied and common tools were used.

The above listing does not exhaust the potential one-hand maintenance actions, but samples

of troubles studied by Bilinski [2] and McKendry *et al.*[4] have shown that the basic types of tools used in almost all cases are those included in the above list. Bilinski's study listed a soldering gun as a frequent tool, but as its use is nearly static, it was not included in the present study.

Apparatus

A photograph of the apparatus employed is shown in Fig. 1. The device was constructed of wood and had a sliding base which allowed for adjustments as required by the different tasks. The first two frames restricted the range of the arm action so that the hand envelope could be described accurately, even under difficult maintenance conditions.

Fig. 1. A photograph of the apparatus employed.

A number of components were mounted on boards which slid into the back of the apparatus. An illustration of this adjustment is shown in Fig. 1 where the subject is in position to turn a socket wrench. Before each subject began an action (such as removing one of the tubes), the board holding the component with which he was to work was moved so that the proper component was directly in front of the center of mass of the subject's hand. A locking nut held the equipment board firmly in place while the action was performed.

When work was completed on one action, the equipment boards were readjusted and re-centered. The entire set of actions required about 15 min per subject. A one-inch white cube was mounted on the same level as the subject's hand. This appeared in all of the photographs and acted as a size reference.

Cameras were positioned at the top and right hand side of the box (see Fig. 1). These took simultaneous slow time exposures (10 sec each) of the action being performed. The resulting effect was a photograph with a blurred outline showing the outer limits of the space envelope required for a specific action.

Subjects

A sample of six subjects was employed. Each of these individuals had hands that were at or above the 95th percentile of the military population [3] in length and breadth. Three of the subjects were engineers who were skilled in the use of tools while the remaining three subjects were individuals who were not especially proficient with the tools required for use with electronic equipment.

By using the mixture of subjects described above, it was possible to get an estimate of the largest and smallest amount of space required for someone with large hands as well as describing a meaningful average for various skill levels. One final point is that all subjects were right-handed; the use of left-handed subjects would cause some slight changes in the data, but should produce no major revisions of the basic trends.

Analysis

By use of a white pencil, outlines of the entire space envelope for each movement were traced on the photographs. There were two photographs taken of each action with one yielding a top view and the other a side view of the space envelope. This was the first subdivision of data for each action.

A series of horizontal and vertical diameters was taken from these views. They started at one-inch from the tips of the fingers and continued for a total of six one-inch intervals over the hand. These measurements provided the second subdivision of data for each action.

A third subdivision was added to more accurately describe motions in which the envelope was not centered directly over the component. For example, when an Allen wrench is employed, all of the action takes place to the side of the component being serviced. This subdivision was obtained by using the center of mass of the component as a reference line which was extended through the space envelope. The diameter measures along the horizontal axis were sub-divided as being either to the right or left of the reference line with "right" and "left" taken in reference to someone facing the target (the component being serviced). The diameter measures along the vertical axis were sub-divided into "up" or "down" using the normal gravitational reference.

Fig. 2 shows the same space envelope with all of the measures described above indicated by a letter of reference. The explanation above the figure explains each of these references.

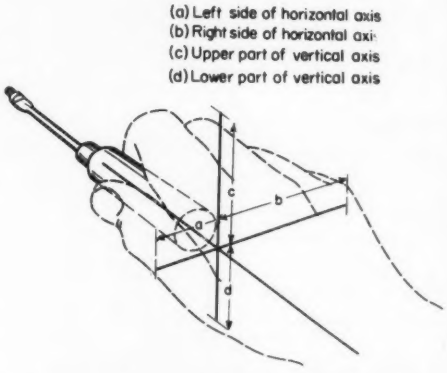

Fig. 2. A three-dimensional representation of the space envelope necessary for a screwdriver action showing all of the measures obtained.

RESULTS

The complete and detailed data obtained for each action will not be reproduced here because of its bulk. They are available in an NTDC report.[1] However, examination of the data showed that the 13 specific actions could be grouped under four basic actions. These resulting envelopes are as follows:

(1) one required to grasp, remove, and replace plug-in units up to $2\frac{1}{16}$ in. in diameter,
(2) one required to turn screwdrivers and spintite wrenches,
(3) one required to grasp, turn and cut with pliers and wirecutters, and
(4) one required to turn wrenches. This has two variations:

(a) where the main turning axis is vertical as was the case with the "Allen" wrench, and

(b) where the main turning axis is horizontal as was the case with the socket-wrench type.

The appropriate data for each of the above envelopes are given in Tables I–V. By reviewing these tables, it is possible to provide estimates on the major factors influencing the size of the space envelope. There seem to be three such factors: (1) the type of motion performed, (2) the size of the plug-in unit being serviced or the size of the tool being used, and (3) the tightness of the socket for plug-in units.

The type of motion was by far the most important factor in determining the size of the envelope required. The horizontal and vertical diameters of the envelope increased 78 per cent to 96 per cent[2] respectively in size from the smallest (the screwdriver action) to the largest (the socket wrench) thereby yielding an average diameter increase of 87 per cent.

The second most important effect which increases the size of the envelope is the size of the component being serviced or the tool being used. This effect has been shown to be less than the one just discussed. For example, while the component diameter increases by 234 per cent from the miniature to the large tube,

TABLE I

Space Envelope for Manipulating Plug-in Units (Tubes of Various Sizes Removed and Replaced). Each Entry is Based on 36 Measurements (6 Subjects × 2 Actions × 3 Tube Sizes)

Distance from end of fingers	*Horizontal axis*				*Vertical axis*			
	Left width		*Right width*		*Up*		*Down*	
	mean	*range*	*mean*	*range*	*mean*	*range*	*mean*	*range*
1 in.	1.19	0.56–2.05	1.69	1.18–2.45	1.23	0.41–1.91	1.31	0.41–2.66
2 in.	1.43	0.68–2.40	2.04	1.41–2.68	1.74	0.75–2.91	1.64	0.83–3.33
3 in.	1.53	0.68–2.40	2.23	1.90–2.56	2.06	1.16–3.25	*1.70*	0.83–3.41
4 in.	*1.59*	0.62–2.35	*2.24*	1.75–2.81	2.30	1.58–3.75	1.55	0.66–3.16
5 in.	1.58	0.56–2.30	2.16	1.37–2.87	2.46	1.50–3.91	1.30	0.41–2.66
6 in.	1.55	0.56–2.30	2.07	0.87–2.75	*2.56*	1.80–3.91	1.08	0.25–2.50

TABLE II

Tools Which Require Hand Rotation (Screw drivers, Spintites). Each Entry is Based on 12 Measurements (6 Subjects × 2 Actions)

Distance from end of fingers	*Horizontal axis*				*Vertical axis*			
	Left width		*Right width*		*Up*		*Down*	
	mean	*range*	*mean*	*range*	*mean*	*range*	*mean*	*range*
1 in.	1.16	0.68–2.00	1.90	1.37–2.50	1.51	0.66–2.25	1.26	0.50–2.08
2 in.	1.45	0.92–2.25	2.31	1.75–2.85	2.00	1.08–2.91	1.62	0.58–2.33
3 in.	*1.49*	0.93–2.25	*2.42*	1.88–2.81	2.26	1.25–3.33	1.67	0.58–2.58
4 in.	1.45	0.65–2.20	2.40	1.75–3.00	2.39	1.25–3.33	*1.52*	0.58–2.50
5 in.	1.41	0.40–1.95	2.32	1.63–2.95	2.31	1.25–3.50	1.36	0.58–2.25
6 in.	1.31	0.35–2.50	2.21	1.68–2.90	*2.44*	1.83–3.58	1.04	0.33–1.83

[2] The percentage calculations given above were computed by subtracting the sum of the diameters of the smaller envelope from those of the larger and expressing the difference as a proportion of the smaller. It should be noted that these values have only relative, not absolute significance. Extrapolations of these data (not just trends) to dissimilar tasks are not recommended.

TABLE III

Space Envelope for Grasping and Turning Tools (Pliers and Wire Cutters). Each Entry is Based on 12 Measurements (6 Subjects × 2 Actions)

Distance from end of fingers	Horizontal axis: Left width mean	Horizontal axis: Left width range	Horizontal axis: Right width mean	Horizontal axis: Right width range	Vertical axis: Up mean	Vertical axis: Up range	Vertical axis: Down mean	Vertical axis: Down range
1 in.	1.64	0.50–2.65	1.67	0.66–2.25	1.20	0.50–1.83	1.69	0.66–3.08
2 in.	1.96	0.68–3.20	2.19	1.18–2.68	1.61	0.66–2.16	2.26	1.00–3.16
3 in.	*2.08*	0.81–3.50	2.44	1.50–3.12	1.72	0.75–2.16	*2.66*	1.00–2.91
4 in.	1.86	0.56–3.40	2.49	1.68–3.37	1.74	0.83–2.41	1.88	0.91–2.41
5 in.	1.49	0.47–3.20	*2.89*	1.68–3.00	*1.83*	1.00–2.50	1.52	0.83–2.08
6 in.	1.29	0.50–2.80	2.36	1.62–2.99	1.83	1.08–2.50	1.40	0.66–2.08

TABLE IV

Space Envelope Required for Using an Average-Sized Socket Wrench to Turn a Nut ($\frac{3}{8}$ in. Base With $3\frac{1}{4}$ in. Shaft). Each Entry is Based on 6 Measures (1 Per Subject)

Distance from end of fingers	Horizontal axis: Left width mean	Horizontal axis: Left width range	Horizontal axis: Right width mean	Horizontal axis: Right width range	Vertical axis: Up mean	Vertical axis: Up range	Vertical axis: Down mean	Vertical axis: Down range
1 in.	2.09	0.47–3.20	2.92	2.25–4.30	2.73	2.00–3.75	2.88	2.08–3.33
2 in.	*2.12*	0.50–3.35	*3.25*	2.37–4.55	2.86	1.75–3.75	2.65	1.66–3.41
3 in.	2.04	0.56–3.31	3.13	2.43–4.35	3.10	1.83–4.25	2.10	1.33–2.50
4 in.	1.86	0.68–3.18	2.94	2.06–4.35	3.23	2.08–4.41	1.65	1.08–2.08
5 in.	1.54	0.56–2.81	2.73	1.93–4.35	*3.25*	2.33–4.08	1.10	0.66–1.75
6 in.	1.31	0.47–2.50	2.55	1.37–4.25	3.11	2.50–4.16	0.76	0.50–1.25

TABLE V

Space Envelope Required for Using an Average-Sized "Allen" Wrench to Remove a Knob (2 in. Length). Each Entry is Based on 6 Measures (1 Per Subject)

Distance from end of fingers	Horizontal axis: Left width mean	Horizontal axis: Left width range	Horizontal axis: Right width mean	Horizontal axis: Right width range	Vertical axis: Up mean	Vertical axis: Up range	Vertical axis: Down mean	Vertical axis: Down range
1 in.	1.02	0.00–1.87	2.95	2.12–4.00	3.36	2.50–4.50	1.79	0.33–5.33
2 in.	1.12	0.00–2.00	3.38	2.12–4.56	3.72	3.00–4.41	2.26	0.66–5.41
3 in.	1.22	0.00–2.13	*3.38*	2.12–4.35	*3.59*	3.00–4.08	*2.54*	0.66–5.08
4 in.	1.31	0.00–2.18	3.07	1.93–3.87	3.31	2.25–3.75	2.50	0.66–4.16
5 in.	1.36	0.00–2.18	2.64	1.62–3.37	2.91	1.25–3.40	2.25	0.66–3.83
6 in.	*1.42*	0.00–2.13	2.31	1.41–3.12	2.77	1.16–3.41	2.01	0.58–3.41

the space envelope diameters increased by an average of only 10 per cent (the vertical diameter increasing 23 per cent and the horizontal diameter decreasing by 3 per cent).

The screwdriver (or spintite) envelope experienced very little increase as a function of size which can be seen from the fact that while the tool size increased by a factor of 290 per cent in length, the diameter of the envelope increased by an average of only 2 per cent (a 2 per cent increase along each diameter). However, it must be remembered that there was no change in the width of the screwdrivers. It is therefore apparent that the tool or component width is the critical dimension in determining the size of the hand envelope resulting from the use of that tool or component. While the effect of increasing tool length is slight (less than 1 per cent increase along each diameter per 100 per cent increase in length), the effect of increasing width is greater (5 per cent average diameter increase per 100 per cent tool diameter increase).

The last important factor in the determination of envelope size was the tightness of the component mounting socket (where applicable). This finding appeared largely as a side effect in the present study. It was found that the average tube had a space envelope almost as large as the large tube. A reinspection of the experimental apparatus showed that the average tube fitted more tightly into its socket than did the other two tubes, which seemed to be equivalent to each other in tightness. This suggests that tightness of the component socket as a variable is almost as important as component size, in determining the dimensions of the space envelope. A rough estimate would gauge the tightness effect at about a 10 per cent increase in size going from the normal to a very tight socket. (The actual increase observed was 5 per cent when the size factor was removed.)

DISCUSSION AND CONCLUSIONS

Certain precautions should be kept in mind when applying the data obtained. It must be remembered that all of the men used in the study had hands at or above the 95th percentile size of the military population. Therefore, the mean and ranges can not be interpreted as they would be in the usual random sample study where the largest source of variance may be due to variations in hand size. Instead, the ranges should be viewed primarily as the product of individual differences in such intangibles as skill, economy of motion and chance.

Considering the above factors, it might at first seem that the data is conservative and that the use of the minimum values obtained would provide adequate space for efficient maintenance actions; but use of these values would not take account of the fact that one individual may require more space on one side of the imaginary center line while another needs more space on the other. For example, subject 1 may have a space envelope which bulges to the right when using a pliers while subject 2 may use more left space for the same action. In essence, this means that if space allowances were based on the minimums a space too small for anyone might result. Therefore, the following interpretations are suggested: (a) design based on the maximums would provide a comfortable space for any maintenance man, (b) design based on the means should provide adequate space for the appropriate maintenance actions, and (c) minimums should be disregarded for maintenance design.

As a second caution to designers, the percentage figures used to illustrate the factors affecting envelope size are rough approximations and, while they are sufficiently accurate to determine the relative order of various factors determining the envelope size, they probably should not be used to estimate the space required for a tool not considered in this study.

The problems raised in the course of this study have exceeded the solutions provided. The sample studied was not large enough or varied enough for delineation of all the factors which control the size of the dynamic hand envelope. Despite these limitations, the data should provide guidance in an area devoid of information; and the photographic technique developed can be recommended as a rapid and

economical method for space envelope studies of this nature.

SUMMARY

In examining present and anticipated maintenance tasks, four basic types of dynamic maintenance actions were isolated: (a) the motion required to turn screwdrivers and spintites, (b) the action required to grasp, remove and replace plug-in units, (c) the action required to grasp, turn and cut with pliers and wirecutters, and (d) the motion required to turn wrenches.

Research by Bilinski[2] and McKendry *et al.*[4] has shown that the actions described above are adequate for the great majority of tool actions required in servicing troubles. The only common motion not listed above is that required in holding a soldering iron. This is a static act, not a dynamic motion; and, therefore, it was not investigated as part of this study.

The space envelope required for each of the above actions was determined by a photographic process. Subjects employed in the study included experienced and inexperienced subjects with hand sizes at or above the 95th percentile for a military population. The dimensions for each envelope were obtained.

Three factors influence the size of the space envelope required for a task. The first, and most important, was the type of hand action required. The second was the size of the component or tool being manipulated, and the third, the tightness of the component mounting socket.

REFERENCES

1. Baker, P., McKendry, J. M. and Grant, G., Design for Maintainability, Technical Supplement III: Anthropometry of One-handed Maintenance Actions. U.S. Naval Training Device Center Report 330–1–3, Port Washington, New York, June, 1959.

2. Bilinski, C. R., Utilization of Hand Tools in U.S. Navy Electronic Equipment Maintenance. U.S. Naval Electronics Laboratory Report No. 888, San Diego, California, Feb., 1959.

3. Jones, C. E., Kobrick, J. L. and Gaydos, H. F., Anthropometric and Biomechanical Characteristics of the Hand. Q.M. Research and Engineering Center Report No. EP 100, Natick, Massachusetts, Sept., 1958.

4. McKendry, J. M., Grant, G. and Corso, J. F., Design for Maintainability; Technical Supplement II: Survey Information (A) Description of the Training Device Technician, and (B) Engineers' Estimates of Circuit and System Maintainability. U.S. Naval Training Device Center Report No. 330–1–2, Port Washington, New York, July, 1959.

Acoustic Noise, An Informational Definition

ALAN A. BURROWS, Life Sciences Group, Douglas Aircraft Company, Inc., El Segundo, California.

Some theoretical considerations are made on the previous definitions of acoustic noise and the classification of its effects on organisms. A definition is proposed relating acoustic noise to its environmental source and the informational content of the specific task in which it occurs. An experimental study explores this proposal.

INTRODUCTION

Acoustic noise has always been ill-defined. Engineers formulating standards[1] in 1942 suggested "Unwanted Sound" as a useful description and Berrien[2] suggests that, "this statement, bearing as it does a heavy value judgment, appears to be widely accepted amongst physicians, engineers and the like". Unfortunately, there has been no indication of how the judgment should be made. Some psychologists have confounded even this barren definition by suggesting that sounds become noise by virtue of their musical content. Dockeray, for example, stated that, "If a complex of vibrations is not in harmonic ratio, we speak of the result as noise".[3] Dashiell's thesis was[4] that, "When air vibrations are non-periodic and irregular or are less than two vibrations, they produce noises rather than tones". These comments may come as a relief to those who love the older composers, but as Kryter points out,[5] "It is difficult to define noise in terms of any of the physical parameters of sound waves, such as complexity of wave form or intensity. For example, both music and pure tones under certain conditions are commonly considered as noise."

A line from the Morals may throw light on the Engineers' attempt. "In all real desire there is some object that is consciously taken as an end."[6] Sound that is unwanted is surely that sound which is unrelated to the process or the end result of the task being carried out. Nearly all acts produce sound. Whether the sound is perceived or not depends upon the threshold and masking conditions at the time. The sound produced by an act shows at least that the act is proceeding or that it is complete. If this is the total information value of the sound, then the sound could be as simple in its nature as the intelligence it is conveying, that is of only a single bit in extent. Any complication of the sound in this case, by increase of bandwidth, intensity or duration over and above the simplest possible pattern to convey one bit of information, is informationally redundant. To go further, we may hypothesize that in the case where a sound conveys only one bit of information about the situation producing it, there is redundancy when the absolute thresholds are significantly exceeded. Greater or more complex transmission over and above these levels is unrelated informationally to the task or process producing the sound and may legitimately be called noise. Thus, we could postulate that unwanted sound or *noise is that auditory stimulus or stimuli bearing no informational relationship to the presence or completion of the immediate task.*

Where sound conveys more than the binary minimum of information about its parent situation, calculation is possible, though not always simple, to show the minimum value of signal for the parameter it is describing. The flight engineer, for example, is interested in the conditions that the engine is going (presence or absence of some sound); its speed (frequency); its smooth running (lack of change of pattern) and its synchronization with other engines (presence or absence of beats). Intensity does not seem to be a prime in this situation, except possibly in interaction with frequency in the speed parameter. The *noisiness* of the engine would then in this case be, to a specific extent, a function of the intensity over and above the level necessary to perceive the sound signals describing the above parameters. To explore this example further, it is apparent that the *value* or the need for the sound signals is complicated

by the presence and quality of other displays of the same information. That is, we should expect the noisiness to be related to the fact that the sound is or is not the prime channel describing the situation. Further, the channel describing the situation producing sound might itself be an auditory channel. In the aircraft case, auditory displays have often been used to describe primary flight and engine parameters.[7] If an auditory display were to be used to describe any or all of the items described above and, at the same time, the engine were producing sound with some signal content, two possibilities exist. First, one sound pattern might reinforce the other and improve the overall display of information. Alternatively, one sound might interfere with the other and reduce the overall efficiency of the display.

CATEGORIES

When we speak of the effects of "noise", it is essential to eliminate the long-standing confusion between the categories of effect. Intense sound can be said to have biological effects upon an organism; at the higher levels, these are characterized by apparently purely physical destructive phenomena within body tissues.[8] At slightly lower levels we see more localized damage, at the ear, together with the generalized central effects alleged by Ward.[9] At lower levels still, local damage is marginal and variable between subjects, but indefinite and unpredictable changes of performance are said to be correlated with certain levels and patterns of sound in specific working situations. We also meet simple communication effects where sound reduces the efficiency of auditory transmission systems by masking.

Studies in the past have tended to treat the performance effects of noise as results of the influence of a single ordered stress of varying energy. It is conceivable that the influence upon the organism is a discontinuous function where different parts and activities are affected with respect to specific energy levels and differences in the frequency, phase and information patterns. If the effect is discontinuous, we perhaps should attempt to classify these levels against the apparent modal differences. At the upper energy levels, purely physical effects override all others, in much the same way that a broken arm can be said to override tiredness in an individual's performance. Here there is no mystery about effects; instinctual avoidance actions are witnessed, distinct pain is felt and unmistakable damage can be observed and measured. At 168 dB (6.0 kc/s), for example, 99 per cent of a sample of *Micrococcus pyogenes* var. *aureus* is destroyed in one hour. *Paramecium caudatum* show progressive deterioration ending in cell rupture. In the pig, 171 dB (6 kc/s) for 3 min result in blistering and erythema, but at 150 dB no damage of this sort is apparent. In man, pain is felt in the middle ear at 140 dB and above, with the threshold increased by 39 dB below 15 c/s: At levels only slightly above 140 dB, man shows innate self-protective responses to sound. Stimulation below the pain threshold, at about 135 dB (1000 c/s) in man, may cause disorientation through labyrinthine excitation. The effect may spread over the auditory band with intensities above this level. Symptoms of nausea, vomiting, nystagmus, shifting of the visual field, feelings of forced movements and staggering and falling may occur in susceptible individuals.[8]

At the slightly lower levels, there is less gross physical masking of the more subtle effects. In the narcotized animal, "overall sound levels of 133–137 dB cause stimulation of the reticular activating system of the brain stem as the result of bombardment presumably entering the central nervous system over the eighth nerve". In the human, EEG changes begin at 135 dB overall and show alpha blockade and general cortical desynchronization. Some effects on the deep tendon reflexes have been noted in some at these levels and have been attributed to the same channels.[9]

Ward here suggests that reticular stimulation at these levels may "mobilize the adrenal stress system, and account for some observed changes in muscle tone, inco-ordination, difficulties in

ocular orientation . . . and possible errors in judgment".

Hallowell Davis goes on to categorize the effects in order of descending sound intensity.[10] With long exposure, at all the above levels, we meet permanent hearing loss, then temporary loss and chronic fatigue (where the body reacts, "through the nervous system and ultimately through certain hormones to mobilize its defenses . . . and pays a price ultimately for repeated mobilizations . . ."). As the intensities drop, the effects become difficult to isolate and reproduce, where, "the workman must try harder and be more vigilant to maintain his normal performance". Further, there are the effects of short, unusual or changing sounds which may only vaguely be described as "disturbing or annoying". Davis also describes the effect of what can be extremely low intensity sounds; these can evoke "fear reaction" and contain information at a binary level: "Basically, does it mean friend or foe?"

There seems to be indication that effects of sound fields could be grouped into three self-inclusive categories, basically, as a function of intensity. The highest intensities result in physical, neurological and informational effects; lower intensities give generally information and neurological correlates and the lowest levels may show isolated information effects.

The definition which is proposed concerns itself primarily with description at the informational level, the degree of stimulus where the sheer influence of energy upon tissue has not yet been felt. Effects upon performance might at this level, be said to be due to *noise*, in the informational sense, entering through the auditory channel. Changes in performance could be said to be due to direct interference with the necessary minimum transfer of data for the task, and in fact, may be described as changes in the signal–noise ratio of the control/display/controller information flow. Generally, if the pattern becomes more complex, the ratio will increase. Where the *intensity* increases, however, the informational noise need not always increase, *except as a secondary effect* in that the organism is specifically stimulated and the gross responses themselves generate interference.

A PRELIMINARY EXPERIMENT

An experiment by the author [11] was designed to explore the differential effects of low-level sound on performance in consideration of its information content. Subjects were required to carry out a simple manual tracking activity using a stylus on a flat, horizontal zig-zag pathway. They were instructed to aim for maximum accuracy in following the path, and for minimum time in completing it. If the path side were touched, an error was scored. A further score was the total time taken to complete the task. Touching the side also electrically stimulated an acoustic circuit which selected one of two conditions:

A. A persistent sound field was switched off for a period of 0.5 sec.
B. A sound field was switched on for the same period.

The task was also carried out in quiet, in continuous sound and in sound switched on and off in a random fashion. Sound levels were varied at four levels between 70 and 100 dB re. 0.0002 dynes/cm^2 and were in all cases at 1000 c/s, reproduced by offset generators. The room used was at all times 30 dB down on the lowest level.

The confounded, factorial design yielded several significant conclusions for variance analysis and mean comparison over the twelve male, student subjects and 432 runs:

(a). A significant increase in the error score of the trials in sound compared with the last four quiet runs ($0.01 > P > 0.001$).

(b). A significant decrease in the time taken in the trials in sound compared with the last four quiet runs ($P = 0.001$).

(c). A distinct change in the learning trend with the introduction of sound fields.

(d). The task in intermittent sound took longer than in continuous sound ($0.05 > P > 0.02$).

(e). The task in continuous sound resulted in more errors than when the sound was switched off by error ($P = 0.001$).

(f). The task in continuous sound produced more error than when the sound was switched on *or* off by error ($0.05 > P > 0.02$).

(g). There was *no* difference between time or error performance in any condition with change in intensity.

The mean order of difference for (a) and (b) above was 10.2 per cent of the errors and 7.1 per cent of the time for all subjects. In (d) the mean difference was 10.5 per cent in total time and in (e) errors were different by 16.6 per cent. In (f) the overall mean differences both exceeded 12 per cent.

Another marked effect was the apparent increase of individual differences when proceeding from the initial quiet runs to those in which some sound field was present. In the error score, performances widened in range from the learning band of 11 units of mean error to well over 15 units. The learning trend previous to the introduction of sound, further, was reversed. This had been a steady widening of error score band between subjects at a rate of 0.55 units/run. The introduction of sound coincided with an immediate scatter of scores between subjects and a subsequent decrease of the band at a rate of 1.0 units/run.

Fig. 1 shows the overall performance of the subjects in quiet (learning) and in the sound fields, irrespective of type. The curves are for the mean of all subjects and for the three groups of subjects within the total group, overall runs.

Fig. 2 is a similar breakdown for the time score. This picture shows again the marked effects of the introduction of the sound field to the task, this, though is in contrast to the error effects; the relative position of the subject groups have been reversed, subjects 1–4 for example showing a high error score in contrast to their low time score.

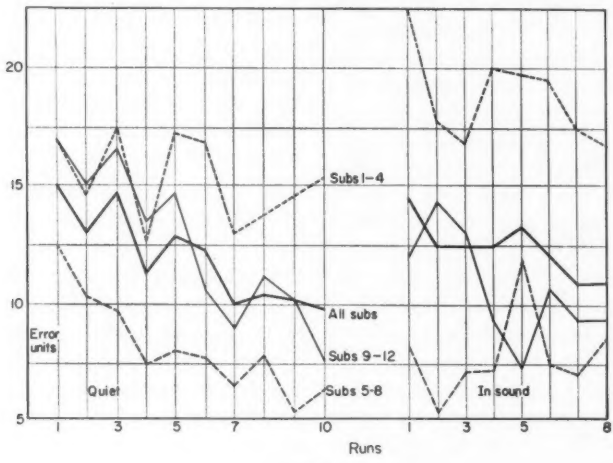

Fig. 1.

To examine the effects of interaction of time and error scores, the ratio T/E was plotted (Fig. 3). The trials in sound were compared with the last four trials in quiet conditions.

The overall mean in sound showed a lower ratio P ($= 0.05$) than in quiet; the sound-on-error lower still ($0.05 > P > 0.02$) and continuous sound even lower ($P = 0.01$), while intermittent sound showed no difference. Sound-off-error showed no effect.

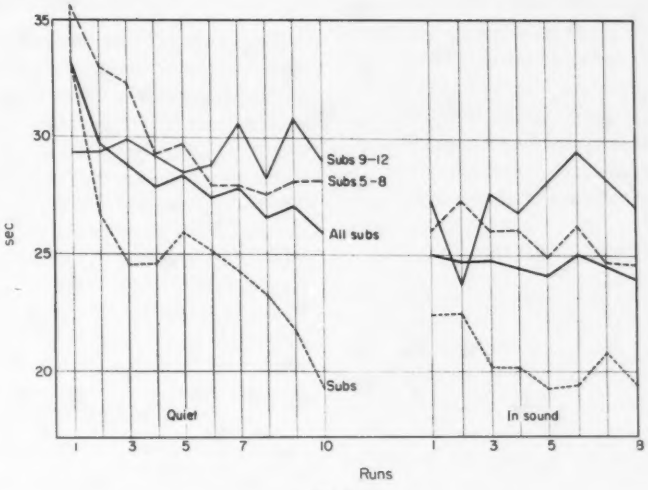

Fig. 2.

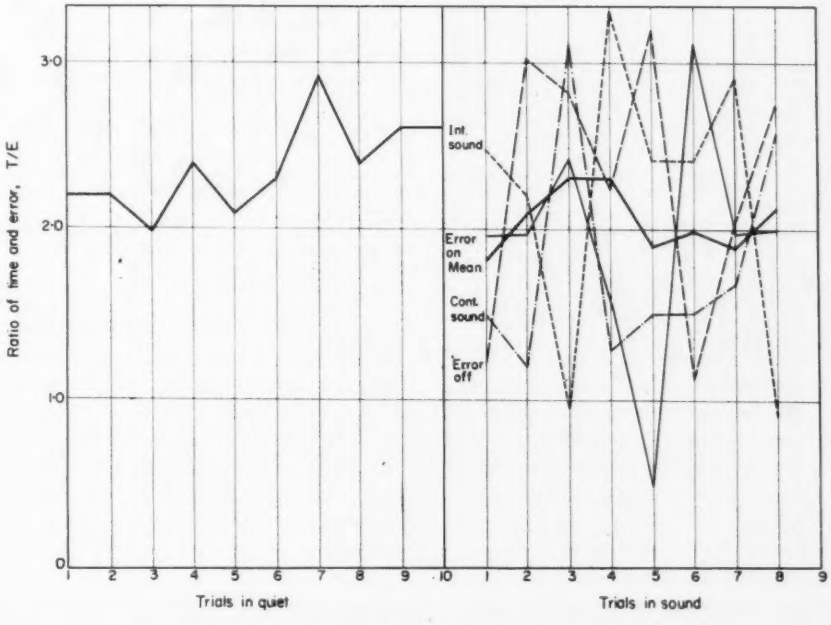

Fig. 3. *T/E* score for sequential trials. All subjects' means, all intensities.

That is, the overall effect of introducing sound to the task was to speed up the rate of working at the expense of error. The two cases where signal was present in the sound had less effect than the overall effect, and very much less effect than continuous sound. Against our hypothesis, intermittent sound seemed not to follow the expected trend.

DISCUSSION

There seems to be little doubt that some disturbance to performance could be attributable to the introduction of sound to the situation. The need for a control situation with initial learning occurring in sound preceding runs in quiet, may be argued; though the importance of the demonstrated internal significances cannot be neglected on this count. That there were changes in performance with sound related to error, was significantly shown. However, the question must be asked, was the sound less noisy and thus did it interfere less than in the unrelated noise state, *or* did the related sound help in the task performance by reinforcing existing visual and tactile error cues? At first sight, there seemed to be little doubt that visual and tactile information was good, complete and could not be grossly improved by the introduction of another modality. However, many studies have shown that this sort of judgment can be misleading.

In consequence, the experiment cannot be taken to be useful, from the point of view of our definition, until the critical experiment is carried out eliminating the possibility of reinforcement of feedback.

CONCLUSION

From the preceding observations and discussion it seems that there may be some value in a re-appraisal of acoustic noise matters from an informational standpoint, and probably it will pay to recall the possibility of discontinuous effects when carrying out performance studies. Further experimentation is required to isolate and explore effects at the lower intensities, where the definition suggested here may have some validity.

REFERENCES

1. *Acoustical Terminology Book.* 224.1, American Standards Association, 1942.
2. Berrien, F. K., Effects of noise. *Psychol. Bull.*, 1946.
3. Dockeray, F., *Psychology.* Prentice-Hall, 1942.
4. Dashiell, J. F., *General Psychology.* Houghton Mifflin, 1937.
5. Kryter, K. D., Monograph Supplement. *J. Speech. Dis.*, 1950.
6. MacKenzie, J. S., *A Manual of Ethics.* p. 3. University Tutorial Press, London, 1948.
7. Forbes, T. W., Auditory signals for instrument flying. *J. Aero Sci.*, **13,** 255, 1946.
8. (a) Allen, C. H., Frings, H. and Roderick, I., The physical effects of high intensity airborne sounds on animals. *J. Cell. Comp. Physiol.*, **31,** 339, 1948.
(b) Allen, C. H., Frings, H. and Roderick, I., Some biological effects of intense high frequency airborne sound. *J. Acoust. Soc. Amer.*, **20,** 62, 1948.
(c) Ackerman, E., Reid, J., Kinsloe, H. and Frings, H., Biological Effects of High Intensity Sound Waves. Report No. 53–82, Wright Air Development Center, 1953.
9. Ward, A. A., Central Nervous System Effects in the Benox Report, ONR Project N.R. 144079, University of Chicago, 1953.
10. Davis, H., in Benox Report.[9]
11. Burrows, A. A., *Noise and Its Relevance to the Working Environment.* Unpublished Thesis, University of London, 1954.

Human Engineering—1911 Style

SHERMAN ROSS and ANDREW A. AINES, American Psychological Association and U.S. Army, Washington, D.C.

A brief historical note is given on a 1911 journal called "HUMAN ENGINEERING" to show an early use of the term. The content and methodology of "human engineering" and related terms as they are used today differ from the 1911 definition.

Typically professions emerge from technologies. In the development of the technology, it is difficult to locate the appropriate name which will isolate and designate the services and functions to be performed. In later stages of development this label becomes of increasing importance since it becomes fixed by specific training requirements, and in some cases by legal restrictions of various kinds.

In the post World War II years in the United States we have seen the emphasis placed on human factors by the military agencies and by some parts of industry. The current interest stems from military problem solving experiences during the war years which involved a variety of military and civilian scientists and engineers with specialized skills in diverse areas such as psychology, physiology, anthropology, etc.

One of the fascinating developments has been the apparent difficulty of finding an appropriate title for these functions. Among the many names suggested have been Human Engineering, Engineering Psychology, Human Factors, Ergonomics and others. It is not our purpose to defend or propose a new title, but to present a brief bit of history on the use of the title, "Human Engineering", during 1911 in the U.S.

During the course of some historical analyses on the development of social concern for the worker in industry, we came across a journal published in Cleveland, Ohio, during 1911 (four numbers) and 1912 (one number).

The editor was Winthrop Talbot, M.D., and according to the masthead Human Engineering was "a magazine for employees and employers published to provide a means of exchanging experiences on the human side of industry and a forum for the discussion of the conservation of human energy". The editor, who could turn a neat phrase, wrote in his first issue almost a half a century ago:

> "The human element in industry is 95 per cent of the industrial problem, whether it be the running of an engine, the construction of a bridge, the mining of ore, the stitching of a shirt or, more complex than all, the making of a home. The psychology of today's industry is relatively in the age of handlooms and tallow candles. 'Human engineering' is a phrase devised by the writer to describe a new profession called into being by new conditions in the industrial world. . . . the mechanics of production in a well organized shop is a marvel of human ingenuity. There is a force of executives—a superintendent, a foreman, the master mechanics, the inspectors—employed at large expense to conduct the mechanical side. But where is the department to look after the men who run the machines? . . . Who is appointed to study the fitness of the operative for his special task, to see that a near-sighted person is not put on work requiring acute vision, that a left-handed man is not running a right-handed machine, that a slow mentality is not expected to do a quick job, that the muscularly weak are not put at laborious tasks? . . .
>
> "In an organization there should be a Department of Human Engineering as contrasted with Mechanical Engineering. . . . Its work is the study of physical and mental bases of efficiency in industry. Its purpose is to promote efficiency, not of machines but men and women, to decrease waste—especially human energy—and to discover and remove causes of avoidable and preventable friction, irritation or injury. The work of the Department is:

1. Condition of labor.
2. Research, experiment and investigation.
3. Modern methods of cooperation."

How science is afflicted with its own fads and fancies is given eloquent voice by Talbot, who wrote:

"Many years of careful research and experiment by eminent scientists have shown that ozone properly applied seems to purify air, destroy decayed organic matter and odors and check the growth of bacteria. This is true when ozone is provided even in light concentrations. The benefits of ozone Purification from a health standpoint and an economic standpoint are also a matter of record. Installations in offices and factories, both in the United States and Europe, have given satisfactory results in that the productiveness of those affected seemed to be increased solely because of improved health conditions. . . . The ozone engineer is the latest specialist of the engineering profession."

In his second issue, published four months later, Talbot writes:

"'Human Engineering' is published not in the pride of possession but from the necessity of acquiring knowledge of men and business matters. Born in a spirit of humility it is now baptized with the waters of chastened exaltation. For in the three months preceding its first issue in January last there had been published in seventy periodicals but thirty articles dealing with the relations of the employer and employed. In the three months following the first issue of 'Human Engineering' we list one hundred and fifty articles—an increase of 500 per cent, due plainly to the emergence of 'Human Engineering' upon the journalistic horizon."

A final passage from the editor deserves inclusion in this discussion, if only because of literary qualities:

"'Human Engineering' is distinctly Melioristic in its outlook. It has not much use for Optimism—it does not believe in painting the lily or refining pure gold. On the other hand, it has less use for Pessimism, as it does not believe in polluting the clear waters of thought with the dyestuffs and outscourings of philosophic mental waste."

Obviously, except for the name of the publication, the claim of origination of the term "human engineering", and the recommendation of research and experiment, Talbot contributes little to the modern concept of human engineering. We believe, however, that some of our colleagues will find this historical morsel of interest.

The Authors

RICHARD G. DOMEY, ROSS A. McFARLAND and ERNEST CHADWICK (*Threshold and Rate of Dark Adaptation as a Function of Age and Time*)

RICHARD G. DOMEY, research associate, Department of Industrial Hygiene, Harvard School of Public Health, Boston, Massachusetts, received his B.S. from Boston University in 1945, the Ed.M. from Harvard University in 1946, and the Ed.D. from the University of Colorado in 1954. His interests are in experimental research, statistics and measurements. From 1946 through 1957, he was a member of the Psychology Department at Northeastern University where he became Associate Professor of Psychology and Associate Research Director. He joined the research staff at Harvard in 1957 where he has been conducting studies in vision, vehicular safety and research of interest to the Department of the Army. He has also cooperated with the Joslin Diabetic Clinic, Boston, Massachusetts, in the analysis of accidents as a function of diabetes. He is a member of APA, KDP and PDK. In addition he is consultant to the Electronics Division, Curtiss–Wright Corporation. Here he has contributed to the development of visual simulation for such aircraft as the 707 Boeing jet, and to the development of simulation of large freight train systems. He is also a consultant for the Sylvania Corporation in regard to the measurement of certain bio-technical aspects of space capsules.

ROSS A. McFARLAND is Professor of Environmental Health and Safety, and Director of the Guggenheim Center for Aviation Health and Safety, at the Harvard School of Public Health. In 1928 he obtained a Ph.D. degree in psychology at Harvard University, and following a year of research on the effects of high altitude on human performance at Cambridge University, England, he taught in the Department of Psychology at Columbia University for ten years. Between 1937 and 1946, he was a member of the Fatigue Laboratory at Harvard University. Since 1947, he has been carrying out research and teaching in the fields of aviation and highway safety, and human engineering at the Harvard School of Public Health. He is well known for his two books, *Human Factors in Air Transport Design* (1946), and *Human Factors in Air Transportation—Occupational Health and Safety* (1953). Since 1951 he has been Director of the Commission on Accidental Trauma of the Armed Forces Epidemiological Board in the Department of Defense which has been responsible for an extensive research program in various universities and research organizations on the problems of industrial and vehicular accidents in the Armed Services. He is a member of the Executive Council of the Human Factors Society.

ERNEST J. CHADWICK received his engineering education at Northeastern University. He was awarded the degree of B.S. in electrical engineering in 1958. During this time he was employed as a co-operative student at Raytheon Manufacturing Company and at the Harvard School of Public Health. He participated in the Northeastern University Co-operative Graduate Program working at the Harvard School of Public Health as a co-operative assignment. He was awarded the degree of M.S. in electrical engineering in 1960. He is a member of Eta Kappa Nu, I.R.E. and the Professional Group on Circuit Theory.

WILLIAM C. STEEDMAN and CHARLES A. BAKER (*Target Size and Visual Recognition*). See page 104, *Human Factors,* May, 1960, Vol. 2, No. 2.

WILLIAM D. HITT and HORACE W. RAY (*A Laboratory Evaluation of the Effects of Electronic Countermeasures on System Performance*)

WILLIAM D. HITT, Assistant Division Chief, is in charge of the Human Factors Section of the Systems Engineering Division of

Battelle Memorial Institute. His major interests are in the areas of industrial and engineering psychology and experimental methodology. Before coming to Battelle, he was an industrial psychologist for Kimberly-Clark, Incorporated. He was a teaching assistant and a research assistant while enrolled as a graduate student at the Ohio State University. Since joining Battelle in 1957, he has designed experiments to evaluate the effectiveness of electronic countermeasures and has conducted research on complex visual displays and on personnel selection procedures. He received his Ph.D. from the Ohio State University in industrial psychology and statistics. He is a member of the American Psychological Association and the Central Ohio Psychological Association.

HORACE W. RAY is a principal psychologist in the Systems Engineering Division of the Battelle Memorial Institute. His major interests are in psychological statistics, experimental design and man–machine systems. His experience includes two years' work with the Air Force Personnel and Training Research Center, Air Research and Development Command as an assistant research psychologist. During this period he focused his attention on problems of personnel selection, training, proficiency measurement and analysis of complex man–machine systems. At Battelle he has concentrated on research in problem solving, electronic countermeasures, design of visual displays and the application of systems-analysis techniques to complex military systems. He received the B.S. degree in industrial psychology from the Pennsylvania State University and his M.A. in psychology from the Ohio State University. He is currently working toward his Ph.D degree in psychology at the Ohio State University.

DAVID B. LEARNER[1] (*A System Measure of Driver Performance*) is Head of the Human Factors Research Group, General Motors Research Laboratory; a position he has occupied since 1958. He is currently concerned with man–machine system research, consumer behavior research and experimental esthetics. He received his B.A. in psychology in 1951, his M.A. in psychology and statistics in 1952, his Ph.D. in psychology and statistics in 1956, all at the Ohio State University. He also served as research assistant and research associate from 1951 to 1953 and was a research fellow at Ohio State from 1955 to 1956. From 1953 to 1955 he was an aviation psychologist in the U.S. Navy (Staff, Chief of Naval Air Advanced Training, Corpus Christie, Texas). Prior to his present assignment at General Motors, he was a research associate at Dunlap and Associates. He is a member of the American Association for the Advancement of Science, the Human Factors Society, the American Psychological Association, the Highway Research Board and the Psychometric Society. He is also Chairman of the Research Committee of the American Society of Mechanical Engineers' Human Factors Group.

[1] Note that this author's affiliation has changed during production of his article.

ALBERT RESNICK and TERRY R. SAVAGE (*A Re-evaluation of Machine-Generated Abstracts*)

ALBERT RESNICK is an associate psychologist with the Psychology Department of the IBM Research Center, Yorktown Heights, New York. Since joining IBM in 1959, he has been concerned with evaluating computer-prepared abstracts and large-scale information dissemination systems. He graduated from the University of Michigan in 1955 and from Wayne State University in 1958. He is an associate member of the Research Society of America.

TERRY SAVAGE is the Manager of the Documentation Analysis Systems Group of IBM's Advanced Systems Development Division. In this capacity he works on problems of information retrieval and language analysis. He joined IBM in 1956 and has worked in the general area of information retrieval since 1957.

He was born in St. Louis, Missouri. He received his A.B. in philosophy and history in 1951, and his M.A. in philosophy, both from Washington University. He has completed his course work for the Ph.D. in philosophy at the University of Pennsylvania. He is a member of the Association for Symbolic Logic and the Association for Computing Machinery. In 1959 he was appointed to the American Standards Association Subcommittee on Machine Coding and Indexing.

H. T. E. HERTZBERG (*Dynamic Anthropometry of Working Positions*) is Chief of the Anthropology Section, Aerospace Medical Division, Wright–Patterson Air Force Base, Ohio. He was born in San Antonio, Texas. He received his A.B. from Rice Institute in 1927 and his M.A. from Harvard in 1942. Between those years he worked in engineering and aerial mapping, graduated from the U.S. Air Corps Flying School at Kelly Field, Texas, in 1929, became a graduate student in anthropology at Harvard and the universities of Texas and Kentucky. After military service in World War II, he went to the Aero Medical Laboratory in 1946. He is author, co-author or editor of more than 40 reports and papers in applied physical anthropology such as "A Prone Position for Pilots", "Anthropometry of Flying Personnel—1950","Annotated Bibliography of Applied Physical Anthropology in Human Engineering" and anthropological chapters in HIAD, HIGED, HIAGSED and other official Air Force Publications. He is a member of the Committee on Physical Anthropology, National Research Council, and is Chairman of the Committee on Anthropometry, Aerospace Medical Panel, AGARD–NATO. He is a member of the American Association of Physical Anthropologists, the American Anthropological Association, the Human Factors Society and other professional groups.

PAUL BAKER, JAMES M. McKENDRY and GEORGE GRANT (*Volumetric Requirements for Hand Tool Usage*)

PAUL T. BAKER is an assistant professor of anthropology at the Pennsylvania State University and serves as a consultant to the Human Factors Group of HRB–Singer, Inc. He obtained his B.A. degree in anthropology from the University of New Mexico in 1951 and his Ph.D. in anthropology from Harvard University in 1956. From 1952 until 1957 Dr. Baker was employed by the U.S. Army Quartermaster Corps where he conducted research in human engineering, clothing design, climatic physiology and human identification. Since coming to the Pennsylvania State University in 1957, he has been teaching and conducting research in physical anthropology. He is a fellow of the American Anthropological Association, and a member of the American Association of Physical Anthropologists and has previously published several articles based upon his research efforts.

JAMES M. McKENDRY is a senior psychologist at HRB–Singer, Inc. where he has been engaged in human factors research in the area of maintainability and troubleshooting. He received his B.S. degree in psychology from the University of Alabama in 1954, an M.S. in industrial psychology from the Pennsylvania State University and is currently completing the requirements for the Ph.D. in industrial psychology at the Pennsylvania State University. At present, he is conducting experimental research on the effects of various packaging techniques upon troubleshooting behavior.

GEORGE GRANT is a senior psychologist and leader of the Human Factors Group, Operations Department, HRB–Singer, Inc. He has been active in basic and applied psychological research since 1955 when he received his B.S. in industrial psychology from the City College of New York. He received an M.S. in industrial psychology from the Pennsylvania State University where he is currently completing work on the Ph.D. in industrial psychology. At HRB–Singer he has been engaged in the

design of equipment for ease of maintenance, operations research analysis of tactical communications systems, design of display systems and task analysis of complex electronic systems. He belongs to the Human Factors Society, the American Psychological Association and the Eastern Psychological Association.

ALAN A. BURROWS (*Acoustic Noise, An Informational Definition*) is Head of the Life Sciences Group within the Equipment and Safety Research Section of the El Segundo Division of the Douglas Aircraft Company, Inc. He served in the British Navy with the Air Radar Facility of the Fleet Air Arm, and spent six years at the Royal Air Force Institute of Aviation Medicine at Farnborough as psychologist and senior psychologist on the permanent Air Ministry staff. He received the M.A. and M.A. (honors) degrees from the University of St. Andrews, and the M.A. and Ph.D. degrees from the University of London in the fields of psychoacoustics and visual displays for high performance aircraft.

SHERMAN ROSS and ANDREW A. AINES (*Human Engineering—1911 Style*)

SHERMAN ROSS is the Executive Officer of the Education and Training Board of the American Psychological Association, Washington, D.C. He recently resigned as Professor of Psychology at the University of Maryland, where he taught and did research in applied and general experimental psychology. He served as a consultant to a number of governmental and private groups, and is currently a member of the Armed Forces—NRC Committee on Bio-Astronautics. His recent research efforts have been in the areas of psychopharmacology, human vigilance and comparative psychology. He is the author of many reports in scientific journals and has contributed to a number of leading books.

LIEUTENANT-COLONEL ANDREW A. AINES is an Army officer on duty in the Office of the Director of Defense Research and Engineering. A member of the American Psychological Association, he maintains an active interest in military psychology and human engineering.

s

ES

er
he
g-
of
d,
nd
ed
al
er
on
rts
gy,
gy.
fic
of

A.
he
nd
an
an
an